Simulation of surface acoustic wave modulation of quantum cascade lasers

Jonathan David Cooper

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
School of Electronic and Electrical Engineering
Institute of microwaves and photonics
September 2013

Acknowledgements

I would like to extend my warmest thanks to many people who have helped shape the course of last four years, without whom I would certainly not have found the resolve to push through and finish this work.

Firstly, I would like to thank my supervisors, John Cunningham, Paul Harrison and Zoran Ikonić, for believing in me in the first place and giving me the opportunity to undertake this project. My thanks are also extended to the Leverhulme trust for funding me throughout my studies. A special mention must also go to my friend and colleague, Alex Valavanis, who has given me countless hours of support and guidance, both in and out of work, without which I surely would not have come as far as I have; I can only hope that through our continued collaborative efforts I get the opportunity to repay the favour.

I could not go any further without extending my gratitude to the long list of people that I've come to know first as colleagues and acquaintances, then later as friends. To all those who I've had the pleasure of occupying *The Quantum Well* with; Viet Dinh, Andrew Grier, Leon Lever, Helen Rafferty and everyone else who has passed through that office; to everyone from the School of Electronic and Electrical Engineering, Andrew Burnett, Craig Evans, Nic Hunter, Andrew Laister, Richard O'Rorke, Mark Rosamond, Chris and Louise Russell (with extra thanks for the hours of finger shredding bouldering), Mike Waite and all the people who I've inevitably missed off; and to all the people who I've ever had the pleasure of meeting during my time in Leeds; thank you for your support, advice and above all your time to make the zeitgeist of the past four years what it has been. I wouldn't change any of it for the world.

I would also like to thank my parents for giving me the best possible start in life; without which I could not have achieved as much as I have. And finally, to my beloved Catherine; thank you for putting up with me during the most trying times of my PhD, that is, writing the majority of this thesis. Your continued support is what has made it all worth while.

To you all, thank you.

Jonathan David Cooper, May 2014

Abstract

Frequency tunable quantum cascade lasers (QCLs) with a broad wavelength tuning range are highly desirable in chemical sensor/spectroscopic applications owing to the wide range of frequencies, from terahertz up to mid-infrared, which must be scanned. QCLs represent the only convenient, low-cost radiation source over this range of frequencies, however little progress has been made in achieving broadband tunability. Surface acoustic waves (SAWs) present an opportunity for achieving broadband modulation by passing a SAW through the gain medium of the QCL. The electric field generated by the SAW via the piezoelectric effect will modulate the carrier concentration within the gain medium causing distributed feedback (DFB) in the QCL. Unlike conventional DFB mechanisms such as etched gratings on the QCL surface, the wavelength of the SAW, and therefore the pitch of the DFB, can be altered allowing tunability in the QCL frequency.

In this work, a theoretical investigation into the interaction between a SAW and the free-carriers within the QCL active region is presented, with particular focus on whether this interaction is strong enough for DFB to occur. Numerical models of both QCL active regions and SAW propagation through semiconductor materials are developed and used in conjunction to simulate the modulating effect of the SAW on the carrier concentration in the QCL. It is shown that the magnitude of this modulation is large enough for DFB occur, giving a DFB coupling constant comparable to, if not larger than, many experimentally demonstrated DFB QCLs. Finally, device design recommendations are presented which aim to maximise this DFB coupling constant in order to give the widest possible tuning range of the QCL emission frequency.

List of publications

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in chapter 2 has appeared in the publication as follows: J.D. Cooper, A. Valavanis, Z. Ikonic, P. Harrison, J.E. Cunningham, "Finite difference method for solving the Schroedinger equation with band nonparabolicity in mid-infrared quantum cascade lasers", J. Appl. Phys. **108**, 1131101 (2010).

The work in chapter 3 has appeared in the publication as follows: J.D. Cooper, A. Valavanis, Z. Ikonic, P. Harrison, J.E. Cunningham, "Stable perfectly-matched-layer boundary conditions for finite-difference time-domain simulation of acoustic waves in piezoelectric crystals", J. Comp. Phys. **253**, 239 (2013).

Journal papers

- Finite difference method for solving the Schrödinger equation with band nonparabolicity in mid-infrared quantum cascade lasers, J.D. Cooper, A. Valavanis, Z. Ikonić, P. Harrison, J.E. Cunningham, *J. Appl. Phys* **108**, 113109 (2010).

- Design of Ge/SiGe quantum-confined Stark effect electroabsorption heterostructures for CMOS compatible photonics, L. Lever, Z. Ikonić, A. Valavanis, J.D. Cooper, R.

Kelsall, *J. Lightwave Technol.* **28**(22), 3272 (2010).

- Transient Analysis of THz-QCL Pulses Using NbN and YBCO Superconducting Detectors, *IEEE T. Terahertz Sci. Tech.* **3**(2), 172 (2013)

- Stable perfectly-matched-layer boundary conditions for finite-difference time-domain simulation of acoustic waves in piezoelectric crystals, J.D. Cooper, A. Valavanis, Z. Ikonić, P. Harrison, J.E. Cunningham, *J. Comp. Phys.* **253**, 239 (2013).

- Comparative study of intersubband absorption in AlGaN/GaN and AlInN/GaN superlattices: Impact of material inhomogeneities, C. Edmunds, L. Tang, M. Cervantes, M. Shirazi-HD, J. Shao, A. Grier, A. Valavanis, J.D. Cooper, D. Li, G. Gardner, D.N. Zakharov, Z. Ikonić, D. Indjin, P. Harrison, M.J. Manfra, O. Malis, *Phys. Rev. B* **88**, 1 (2013).

Conference proceedings

- Intersubband line broadening in near-infrared Al(0.55)GaN(0.45)/GaN quantum wells, A. Grier, J.D. Cooper, A. Valavanis, L. Lever, Z. Ikonić, D. Indjin, P. Harrison, *UK Semiconductors*(2012).

- Modelling surface acoustic wave modulation of the carrier concentration in quantum cascade lasers for broadband tuneability, J.D. Cooper, A. Grier, A. Valavanis, L. Lever, Z. Ikonić, P. Harrison, J.E. Cunningham, *Internation quantum cascade lasers school and workshop*(2012).

- A scattering rate approach to the understanding of absorption line broadening in near-infrared AlGaN/GaN quantum wells, A. Grier, J.D. Cooper, L. Lever, A. Valavanis, Z. Ikonić, D. Indjin, P. Harrison, C. Edmunds, J. Shao, L. Tang, G. Gardner, D. Zakharov, M.J. Manfra, O. Malis, *6th Space Agency-MOD orkshop on wide band gap semiconductors and components*(2012).

- Surface acoustic wave modulation of quantum cascade lasers, J.D. Cooper, Z. Ikonić, J. CunninghamE, P. Harrison, M. Salih, A.G. Davies, E.H. Linfield, *Proceedings of the International Conference on Advanced Optoelectronics and Lasers*, CAOL, 22 (2013).

- Density Matrix Model Applied to GaAs and GaN-based Terahertz Quantum Cascade Lasers, A. Grier, Z. Ikonić, A. Valavanis, J.D. Cooper, D. Indjin, P. Harrison, *International Workshop on Computational Electronics* (2013).

List of symbols

Semiconductor related

ϕ Wavefunction

E Energy

k Wavevector

m^* Effective mass

V Conduction band potential

z Growth direction of the semiconductor

N_d Doping density

τ State lifetime

n State population

χ Atomic susceptibility

γ_{ij} Transition linewidth

T Transition dephasing time

κ Distributed feedback coupling constant

Acoustic wave related

σ Stress tensor

ϵ Strain tensor

$\mathbf{u}$, u_i Displacement vector

C, C_{ijkl} Elastic stiffness constant tensor

C_{ij} Elastic stiffness constant tensor in matrix notation

e, e_{ijk} Piezoelectric tensor

e_{ij} Piezoelectric tensor in matrix notation

ϕ Potential

ρ Charged carrier concentration

$\mathbf{D}$, D_i Electric displacement vector

t Time

Contents

Chapter 1

Introduction

This thesis develops a theoretical model for the simulation of surface acoustic wave (SAW) modulation of quantum cascade lasers (QCLs). The main aim is to establish the level of modulation achievable in order to determine if this is a viable method for achieving a tunable, single mode emission QCL for use as a high power, convenient radiation source for spectroscopic applications. This chapter gives and overview of QCL and SAW technologies as well as the applications for a tunable terahertz and mid-infrared radiation source. It is assumed that the reader is familiar with semiconductor and laser physics and has a basic understanding of simple computational techniques.

1.1 Quantum Cascade Lasers

Since their first demonstration in 1994 by a group at Bell Labs [1], quantum cascade lasers (QCLs) have attracted a great deal of interest from researchers and industry alike. Aside from their potentially high efficiencies and output powers, the most attractive feature of QCLs is the ability to choose the emission wavelength through careful design work, rather than being confined by the range of available bandgaps as is the case with conventional semiconductor lasers. This restriction occurs within conventional semiconductor lasers because the emission is generated by electrons decaying across the bandgap, therefore the emission wavelength will always correspond to the energy gap of the bandgap. The emission from QCLs, on the other hand, is generated from electrons scattering between different electron energy levels which exist within a carefully designed, periodic semiconductor heterostructure. These electron energy levels may be altered by changing the widths of the quantum wells (QWs) which make up the metros truce, allowing the transition energy of the laser, and hence the emission wavelength, to be chosen at will.

QCLs have enabled high powered laser emission at wavelengths which were previously inaccessible to conventional lasers, namely in the far-infrared and terahertz regions of the electromagnetic spectrum. This has lead to a large amount of research interest within these frequency ranges, with potential applications including medical imaging, security screening, spectroscopy for the detection of illicit materials as well as explosives, and communication applications. (For an overview of QCL applications see [2, 3, 4, 5, 6]). Since QCLs are currently the only convenient, high-power source of terahertz and far-infrared radiation, a great deal of research has gone into their development. In particular, there is a strong need for reliable QCLs which can operate up to room temperature and beyond, as this will allow QCL applications to become commercially viable.

Infrared QCLs have far better performance than their terahertz counterparts. This is mainly due to the larger energy gap between the upper and lower lasing levels meaning that thermal back filling will be significantly reduced, resulting in higher operating temperatures, and the larger wavelengths meaning that sub-wavelength waveguides need not be used to couple the generated radiation out of the laser meaning lower loses. This higher performance is reflected in the important dates in the development of infrared QCLs. The first room temperature operations (pulsed-mode) was achieved in 1996 by Faist *et al.* [7], with the continuous wave room temperature operation being realised in 2003 by Yu *et al.* [8]. By comparison, terahertz QCLs have still yet to demonstrated at room temperature with the highest achieved temperature to date being 199.5 K achieved in 2012 by Fathololoumi *et al.* [9].

In the simplest explanation, the operation of QCL is analogous to a ball rolling down a flight of stairs, where the ball represents the flow of electrons through the structure, and the stairs represent the energy levels which exist within the QCL. Firstly, electrons must be transported along the flat section of each *step*, a process which is facilitated by the applied bias across the QCL forcing electrons from one side to the other. Once at the *drop* between each step, electrons will scatter to the lower energy level emitting photons as they do so. This is, of course, a greatly over simplified picture of a QCL. In reality, their periodic structure is made up of two different alternating layers; an active region and an injector. The active region is made up of an upper and lower laser level, and it is the transition of electrons between these two states that produces photons. The upper laser level is designed to have a relatively long lifetime, such that electrons will will exist in this state for a relatively long time before they decay. This, along with the fact that the lower laser level has a short lifetime due the large overlap with an extraction state which moves electrons very quickly into the

injector region, means that there is a population inversion between the two states which is a key condition for laser action. The injector has two main purposes; to couple strongly to the lower laser level to quickly transport electrons out of the active region (hence giving the lower laser level its short lifetime), and to quickly transport electrons to the upper laser level of the next active region where they can decay via stimulated emission. The two main types of QCL are defined by the mechanism which is employed to scatter electrons through the injector region. Bound-to-continuum designs use a miniband of strongly coupled states which are also strongly coupled to the lower laser level, making it preferable for electrons to tunnel out of the bottom of the active region and through the miniband.The upper laser level of the next active region normally extends into previous miniband, such that electrons at the bottom of the miniband will tunnel straight into the next active region. Resonant-phonon QCLs separate the lower laser level of one active region and the upper laser level of the next active region by LO-phonon energy, making electrons in the lower laser level scatter very readily into the next active region by emitting an LO-phonon.However, because the energy of the lasing transition is relatively small compared to the LO-phonon energy, scattering via phonon emission cannot happen too close in real space to the lasing transition, as it will become preferable for electrons in the upper laser level to scatter via phonon emission and therefore affect the population inversion. Instead, electrons in the lower laser level are forced to tunnel into adjacent QWs via an almost degenerate state which extends into the injector, where they can safely scatter via phonon emission.

1.2 Surface Acoustic Waves

The phenomenon of acoustic waves bound to a propagation mode close to the surface of an elastic solid was first explained by Lord Rayleigh in 1885 [10]. Surface acoustic waves (SAWs) would remain as little more than a mathematical concept until 1965 when the invention of the interdigitated transducer (IDT) by White and Voltmer [11] allowed SAWs to be easily generated on piezoelectric substrates, enabling possible applications of SAW technologies to be explored.

An IDT device consists of a series of perpendicular rectangular strips or fingers patterned onto a piezoelectric substrate as shown in figure 1.1, with alternate fingers connected to opposite supply rails which, in tern, are connected to opposite sides of an oscillating voltage source. Since alternate fingers are connected to opposite sides of the voltage source there will be an electric field of alternating sign between each pair of fingers. This electric field

will permeate into the substrate where, owing to the piezoelectric effect, it will generate areas of localised strain in between each finger pair. Furthermore, because the oscillating applied voltage source, these localised areas of strain will oscillate in magnitude causing an acoustic wave to propagate radially downwards from between each finger pair, as shown by the dashed lines in figure 1.2. As there are multiple finger pairs, multiple acoustic waves will be generated within the substrate and will combine to give an interference pattern which will be dependent upon the spacing between each finger. For an IDT device the spacing from the middle of one finger to the middle of the next is half the desired wavelength of the SAW, $\frac{\lambda}{2}$, which is determined by the acoustic velocity of the substrate. With this spacing between fingers, and therefore between the epicentres of each generated radial acoustic wave, the part of the acoustic wave which travels along the substrate's surface will constructively interfere with the other acoustic waves generated as it passes through their points of generation, as also shown in figure 1.2 by the red and blue shaded areas within the substrate. This constructive interference gives a large acoustic wave magnitude near the substrate surface, whereas destructive interference gives a small magnitude at greater depths. Since the spacing between each finger is physically set by the desired SAW wavelength, the IDT will have a resonance at this frequency and applying a frequency at anything other than this frequency will produce less constructive interference and therefore a less powerful SAW. The shape of the response of an IDT follows that of a cardinal sine function, owing to the points of constructive interference coming in and out of alignment as the applied frequency is varied, as discussed further in chapter 4.

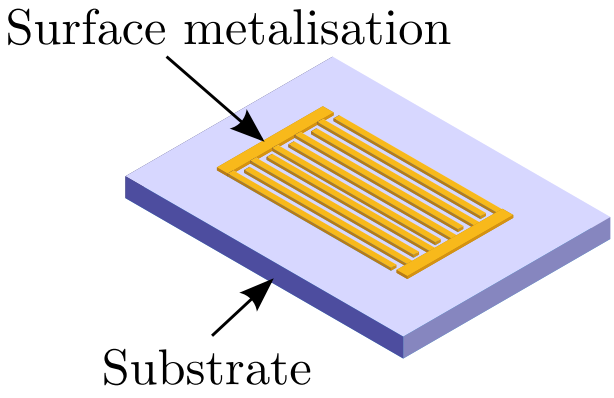

Figure 1.1: An IDT device made of surface metallised fingers, each connected to alternate bond-pads.

As a greater number of finger pairs will lead to more constructively interfering acoustic

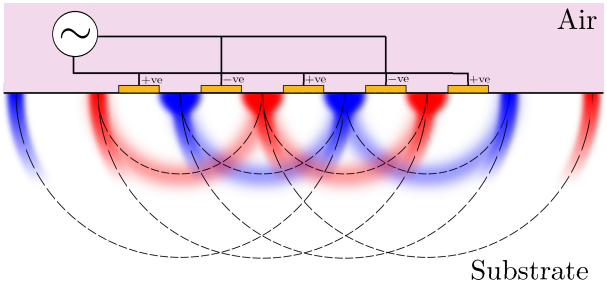

Figure 1.2: Cross-section of the interference pattern of the acoustic waves below the substrate surface, underneath an IDT. The red and blue areas indicate regions of induced stress of different signs which propagate radially downwards from in between each finger pair. (The peak amplitudes of the radially propagating waves are shown by dashed lines for clarity.) With the periodicity of the fingers at $\frac{\lambda}{2}$ of the applied oscillating frequency (as shown here), each peak will constructively interfere with the peak generated underneath the adjacent IDT pair close to the surface. This produces an almost entirely surface bound acoustic wave which propagates from each end of the IDT.

waves and therefore a greater SAW amplitude, IDT devices often employ a large number of finger pairs as is feasibly possible. For the lowest frequency IDT devices, with frequencies of the order of megahertz the SAW wavelength will be of the order of centimetres (since the acoustic velocity in most crystalline solids is of the order of $3000\,\mathrm{ms}^{-1}$), meaning that space is at a premium and the IDT devices may only 5-10 finger pairs [12]. For the highest frequencies achievable — the highest of which was 23.5 GHz [13][1], the SAW wavelength is of the order of 100 nm meaning that space is not an issue and the IDTs typically have up to several hundred finger pairs to produce a more powerful SAW.

An IDT device may be used to convert the energy of SAW passing underneath it into an electrical signal just as it can be used to generate a SAW from a signal — a SAW passing underneath the IDT will induce a voltage in the surface metallisation from the piezoelectric effect; if the SAW is of the same wavelength as the resonant wavelength of the IDT then the voltage of alternate fingers will constructively add. Because of this *sensing* ability of IDTs, a basic SAW device can be made by placing an IDT at each end of the substrate, such that a SAW may be excited by one IDT and detected by the other.

The large research interest and quick development of SAW devices after the invention

[1]Note that this frequency is the highest achieved because of the extreme difficulty in fabricating an IDT on this scale, with fingers that are tens of nanometres across. For a discussion of these fabrication difficulties see [13, 14].

of the IDT meant that by the late 1970's SAW filter electronics were included in a range consumer, commercial, and military devices from TV's to radar systems [15]. By the 1980's, the high insertion losses (> 6 dB) associated with IDTs at the time meant that SAW devices were unsuitable for higher frequency applications such as RF. This led to more exotic designs of IDT and the creation of *pseudo*-SAW devices, which utilise propagation modes that are similar but distinctly different from the original surface bound mode [12]. The development of these pseudo-SAWs, as well as the discovery of new piezoelectric materials, led to a vast improvement in device performance, and by the late 1990's SAW-based RF devices could be found in a wide range of mobile and wireless communication products.

In recent years, a large proportion of the research that utilises SAWs has been focused on quantum transport. The transport of single electrons across quasi-one-dimensional channel (Q1DC), created by depleting a two-dimensional electron gas (2DEG) so that only a thin channel in the centre remains, was first demonstrated by Shilton *et al.* in 1996 [16]. Since then, a wealth of research has gone into exploring how SAWs interact with 2DEGs. There have been many theory papers published to quantitatively explain this phenomenon [17, 18, 19, 20], all of which rely on the exact solution of the differential equations which describe SAWs [21]. While there are many very useful applications for single electron transport such as quantum information processing, the impact of this has been small because it can only be done within an experimental setup (i.e. at very low temperatures) and therefore most useful applications have yet to be properly realised. Such experiments have, however, allowed for great developments in the emerging field of quantum metrology, whereby physical parameters may be measured incredibly accurately using quantum effects.

In terms of numerical simulation work, there has been little done on SAWs propagating through crystalline semiconductors. In the literature, Cambiaggio *et al.* published a paper in 1978 defining a scheme for converting the differential SAW equations into a set of finite difference equations [22]. While this scheme was specific for the $LiNbO_3$ material system, and therefore lacks generality, it is a good starting point showing how the finite difference method can be applied to SAWs. There have also been other more general simulations of SAWs such as [23], however these often employ very complicated methods such Greens functions making them difficult to implement.

1.3 Surface acoustic wave modulation of quantum cascade lasers

The concept of utilising a SAW to induce distributed-feedback-like modulation of a QCL was proposed in 2003 by Kisin *et al.* [24]. The idea was to generate a propagating SAW within the QCL ridge such that the electric field wave that moves with the SAW (via the piezoelectric effect) will interact with the active region, attracting free carriers to the areas of high potential, therefore modulating the carrier concentration along the axis of the SAW propagation direction. This modulation in carrier concentration will affect not only gain of the QCL active region, but also the electrical contribution to the refractive index, i.e. the atomic susceptibility [25], giving them both a periodic modulation corresponding to the periodicity of the SAW. It is the combination of these two effects that preferentially amplify the electromagnetic modes within the laser cavity whose points of maximum amplitude (dependent upon the refractive index) correspond with the points of maximum gain; the same mechanism which distributed feedback (DFB) QCLs utilise [26]. Since the modes which exist within the laser cavity vary with frequency, i.e. the Fabry-Pérot modes of a laser cavity, this mechanism allows one mode to selected thereby achieving single mode emission from a QCL. The advantage of utilising a SAW-modulated QCL approach to achieve this is that the frequency of the SAW may be altered, changing the periodicity of the modulation and the frequency of the selected mode therefore allowing tunability of the single mode emission, unlike DFB QCLs which achieve single mode emission via an etched grating in the top contact of the QCL giving a fixed emission frequency.

A tunable, single-mode QCL is highly desirable in chemical sensor/spectroscopic applications where a wide range of frequencies, from terahertz up to mid-infrared, are required to excite different inter- and intra-molecular vibrational modes within a sample to give information about its chemical composition and structure. (For a overview of these applications see [27, 28, 29, 30, 31] and the papers therein.) Unlike current spectroscopic sources such as lead-salt lasers [32], systems which utilise a high-power gas laser to excite an Austin switch and generate a broadband pulse over the desired frequencies [33], or external cavity QCLs [34], SAW-QCL devices represent the only high-powered, convenient sources with the potential to be tuned over the required range of frequencies [35].

Despite the potential advantages of SAW-modulated QCLs, the only attempt within the literature to physically realise a device of this nature was by Salih *et al.* [36]. Within their design, a QCL ridge was fabricated and mounted onto a separately fabricated SAW device, in

between the two IDTs in the path of the SAW such that the long axis of the QCL is aligned with the SAW propagation direction, as shown in figure 1.3. Since the SAW will be generated within the substrate which the QCL is mounted on top of, this design relies upon the SAW moving from the substrate up into the QCL ridge in order for modulation of the QCL active region carrier concentration to occur. The potential SAW propagation paths are illustrated in figure 1.4. There are two features of this design which aim to facilitate this movement of SAW energy into the QCL ridge. Firstly the facets of the QCL angled producing a sloped surface which the SAW must travel up. This aims to reduce the scattering of the SAW at the discontinuity between the surface of the substrate and the QCL ridge, analogous to a tapered waveguide, such that SAW will travel smoothly form the substrate into the QCL ridge. This does, however, reduce the effectiveness of the QCL ridge as a resonant cavity and therefore degrades the performance of the QCL; full details of which can be found in [37]. Secondly the height of the QCL ridge has been reduced, from around 15 μm typically used with this QCL design[2] down to 5 μm, as detailed in [39]. This aims to reduce the size of the discontinuity between the substrate surface and the QCL ridge, again reducing the amount of scattering of the SAW at thus continuity and increasing the amount of SAW energy which enters the QCL ridge. Despite these design features, a measurable modulation in the QCL output from the SAW was not detected with this device. This was in part due to the difficult nature of the fabrication of this device (full details of which can be found in [36]), but also due to the unknown amount of SAW energy which is scattered at the discontinuity in the surface of the device and therefore the unknown level of modulation which can be expected in the QCL active region carrier concentration.

With these concepts in mind, a theoretical study of SAW propagation through QCLs is required to answer three important questions about these devices that cannot be easily answered experimentally. Namely, the amount of SAW energy which moves from the substrate of the SAW device up into the QCL ridge, the level of modulation of the carrier concentration within the active region of the QCL (as well as the depth of that modulation since the presence of free charge may screen the effect of SAW and stop it modulating the entire depth of the active region), and how that modulation in carrier concentration effects the gain and refractive index of the QCL. These questions are the concern of this thesis.

[2]The QCL active region used by Salih *et al.* was the Luo *et al.* three-well active region design [38].

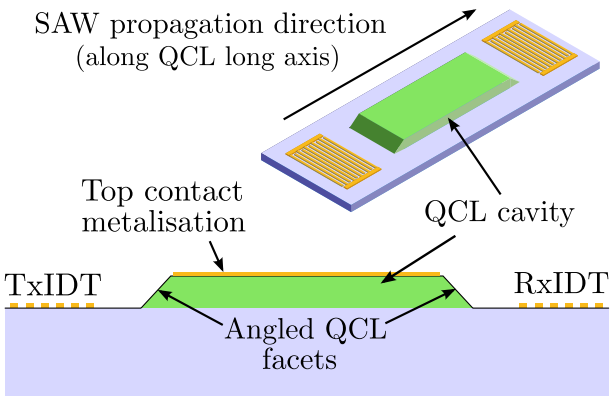

Figure 1.3: The SAW modulated QCL device design as proposed by Salih *et al.* [36] consisting of a QCL cavity mounted on top of SAW device. The propagation of direction of the SAW is aligned within the QCL long-axis and the QCL facets have been angled to try and help the SAW move from the substrate into the QCL cavity.

1.4 Thesis Structure

The main of this thesis is to provide the tools to perform theoretical study into SAW propagation through QCLs in order to test the validity of using SAWs to achieve DFB modulation of QCLs. There are three main unknowns which this study aims to find, as they cannot be easily determined experimentally. Namely, the amount of SAW energy which moves from the substrate of the SAW device up into the QCL ridge, the level of modulation of the carrier concentration within the active region of the QCL (as well as the depth of that modulation since the presence of free charge may screen the effect of SAW and stop it modulating the entire depth of the active region), and how that modulation in carrier concentration effects the gain and refractive index of the QCL.

Chapter 2 is concerned with the modelling of semiconductor heterostructures, which is required to determine the electronic structure of QCLs. A novel method for the direct solution of the nonparabolic Schrödinger equation is also derived, allowing the quantum states within a heterostructure to be found both accurately and reliably in material with high conduction band offsets such as GaN. The self-consistent Schrödinger-Poisson solution is also examined for both performance and accuracy, allowing charge effects to be included in the calculation.

Chapter 3 derives the acoustic wave equations of motion from first principles and details how these may be discretised to produce a time-dependent model of acoustic wave propagation. Since this is an open domain simulation, novel perfectly-matched-layer boundary

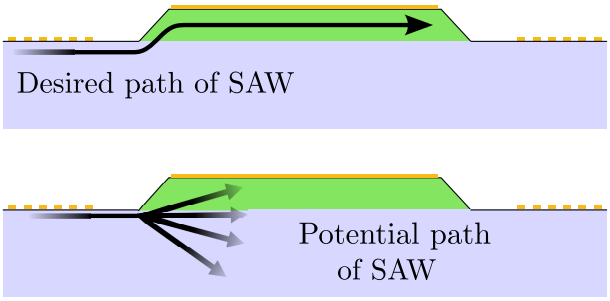

Figure 1.4: The potential paths of the SAW as it moves from the substrate into the QCL cavity, either moving completely up into the QCL cavity and staying as surface bound propagation mode *(top)* or scattering in to a bulk propagation mode as it hits the discontinuity in the surface such that much of the energy of the acoustic wave does not enter the QCL cavity and does not contribute to modulation *(bottom)*.

conditions are derived to solve the problem of reflections off the simulation domain boundaries interfering the physics within the region of interest. This model is then extended in chapter 4 to include a surface boundary condition such that SAW propagation may be simulated. A model of IDTs is also developed to excite the SAW simulation and an experimental validation of this model is also provided.

Extensions to the SAW model, which are required to simulate SAW modulation of QCLs, are presented in chapter 5 along with a rate-equation model of QCL active regions which is used to determine the gain and total refractive index of the QCL. These two models are then combined to show the level of modulation in the gain and refractive index which can be expected from SAW modulation. These results, as well as those from previous chapters are summarised in chapter 6, along proposals for further work for which this thesis could serve as a basis.

Chapter 2

Modelling semiconductor heterostructures

Semiconductor heterostructure devices, such as QCLs, utilise careful engineering of the electronic band structure in order to enhance desirable properties required by the device. This *band engineering* is achieved by controlling the thickness of each individual semiconductor layer within the heterostructure, and therefore the thickness of each quantum well / barrier within the devices band structure. Within the context of QCLs[1], the desirable properties required are those which allow the device to be used as a laser gain medium, i.e. a population inversion[40]. Since these properties depend upon the quantum processes (predominantly scattering and tunnelling within QCLs) through which free carriers move through the band structure [41], it is required that the quantum states, on which these properties depend, are accurately found. This is done by solving the Schrödinger equation.

Since the heterostructure devices utilise confinement in one direction, the only variation in their structure is along the direction of this confinement meaning that the quantum states within the structure need only be found in one dimension. Furthermore, as QCLs are steady-state devices, the time-dependent variation in these states may be ignored and they may be found using a time-invariant method. Therefore, the time-independent one-dimensional Schrödinger equation may be used to find the desired quantum states. While there are freely available programs which will solve Schrödingers equation such as Nextnano [42], using third-party software is less flexible and not open to further development/improvement in the way that *in-house* software is. Furthermore, within the quantum electronic group at the

[1]Since holes do not contribute to the operation of QCL devices, the scope of the discussion will limit to the consideration of holes for the remainder of this thesis.

University of Leeds there is a long history of developing in-house simulation tools including a solution to the time-independent one-dimensional Schrödinger equation, details of which can be found in [41, 43, 44, 45] and the references therein.

Within this chapter a model for solving the time-independent one-dimensional Schrödinger equation is presented allowing the quantum states within a semiconductor heterostructure to be accurately determined. This model is then extended to use a nonparabolic dispersion relation which increase the range of validity of the model, allowing heterostructures with large conduction band discontinuities to still be simulated accurately. Charge effects on the semiconductor bandstructure are considered next by solving the Poisson equation for the distribution of free charge within the structure as found by solving the Schrödinger equation. By combining the Schrödinger and Poisson solutions into one self-consistent loop, as shown at the end of this chapter, the quantum states within a heterostructure may be found including the effect free charge which changes the shape of the semiconductor bandstrucutre (and in tern the free-charge distribution). This allows QCL bandstrucutres to be accurately modelled with different doping concentrations, and therefore different amount of free charge within the active region, which is crucial in determining the effect of the SAWs modulation on the QCL active region carrier concentration on the operations of the QCL. (This is discussed in detail in chapter 5.)

2.1 Time-independent Schrödinger equation

The simulation of optoelectronic semiconductor heterostructure devices often begins with solution of the time-independent SE:

$$-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\left(\frac{1}{m(E,z)}\frac{\partial}{\partial z}\psi(z)\right)+V(z)\psi(z)=E\psi(z) \tag{2.1}$$

where m is the mass of a particle described by the wavefucntion, ψ, confined within a potential, V, with some energy, E. The solution of this equation gives a set of basis states (eigenvectors) and corresponding energies (eigenvalues) from which one can predict the carrier distribution as well as their transport between states with a heterostructure [41].

2.1.1 The effective mass approximation

The effective mass approximation (EMA) is an essential tool for solving the Schrödinger equation (SE) in semiconductor heterostructure devices. As well as greatly reducing the computational time required to find a solution to just a fraction of a second,[2] this approx-

[2] A full bandstructure calculation can take many hours to complete.

imation simplifies the problem immensely allowing one to have an appreciation for what is going on within the electronic structure of a semiconductor, without the need to implement a complicated, in-depth model of the underlying physics (such as density-functional-theory simulation). The main assumption is that charge carriers confined within a semiconductor (i.e. electrons and holes) will have a relatively low momentum, and therefore on the curve of energy vs momentum, the dispersion relation, only relatively low energies need to be considered, i.e. much lower than the band gap of the semiconductor. For low momenta the dispersion relation in most semiconductors is approximately parabolic, which is the same as the dispersion relation for a particle in free space. An empirical scaling parameter called the *effective mass* can be used to fit the parabolic dispersion relation of a semiconductor to that of free space, hence treating charge carriers within a semiconductor as if they are in free space, only with some new *effective* mass.

Assuming that the dispersion relation for charge carriers is parabolic inherently implies another important approximation; the *envelope function*. The best way of explaining the envelope function is to carefully consider why it is implied by the EMA. Firstly, consider what the EMA is actually doing to the way charge carriers are treated; they are now thought of as behaving exactly like they would do in free space only with a different value for their mass. Secondly, consider the difference in electrical potential energy between free space and in a semiconductor; free space will have a constant potential (depending on the applied electric field), whereas a semiconductor has a rapidly varying periodic crystal potential, created by the outer shell of electrons surrounding each atom. Because charge carriers within a semiconductor are now assumed to behave exactly as they would in free space, the electrical potential within the semiconductor must also be assumed to be the same as in free space for the EMA to be valid. This leads to the widely accepted simplification of bulk semiconductors having a valence and conduction band each with constant potential edges, with different materials having different values for the band edges due to their unique crystal potentials. This allows a semiconductor heterostructure to be thought of as a series of one dimensional quantum wells which arise from the discontinuity in band edges between each material. Solving the SE for this simplified quantum well model of a heterostructure does not give the *real* wavefunction of a carrier within the crystal, which is a rapidly varying function on the scale of the lattice constant, rather it gives the envelope function of the wavefunction, which is slowly varying due to the fact the rapidly varying component of the potential is not include in its calculation. This is depicted graphically in Fig. 2.1.

Using this envelope function approximation along with the EMA does not effect the

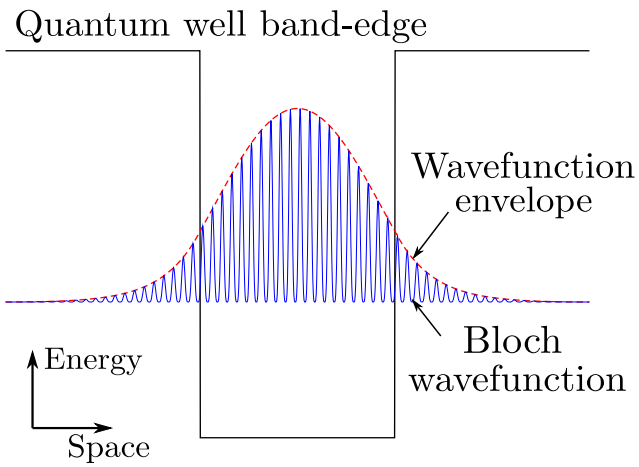

Figure 2.1: Envelope function of the ground within a qunatum well. The variation in energy of the conduction band edge is shown in black with the rapidly varying Bloch wavefunction shown in blue. The envelope function is shown in red.

validity of the quantum mechanics at work within any system simulated. After all, the SE is still being solved for carriers under some kind of quantum confinement. Instead, these are approximations of the material parameters of the semiconductor, specifically that carriers behave as they in free space only with a different mass and some externally applied potential field. This means that these approximations will still produce results that are representative of real systems.

2.1.2 Parabolic dispersion

For a parabolic dispersion the effective mass is constant in energy (although varies spatially as it is different within different materials) and has the form

$$m^* = m_{\text{eff.}}(z)m_0 \tag{2.2}$$

where m_0 is the free space space of an electron at rest and $m_{\text{eff.}}$ is the effective mass coefficient — an empirical fitting parameter which is dependent upon the composition of the semiconductor and therefore spatially varying within a semiconductor heterostructure. See appendix C for the forms of $m_{\text{eff.}}$ for all the material systems examined within this thesis.

Inserting the parabolic form of the effective mass into the time-independent SE (2.1) gives

$$-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\left(\frac{1}{m(z)}\frac{\partial}{\partial z}\psi(z)\right)+V(z)\psi(z)=E\psi(z). \tag{2.3}$$

To solve (2.3) numerically the finite difference approximation (FDA) [41] may be used where

$$\frac{\partial f(z)}{\partial z}\approx\left.\frac{f_{i+1}-f_{i-1}}{\delta z}\right|_{i=z}, \tag{2.4}$$

Applying first to the inner most differential gives

$$-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\left(\frac{1}{m(z)}\frac{\psi_{i+1}-\psi_{i-1}}{2\delta z}\right)+V(z)\psi(z)=E\psi(z), \tag{2.5}$$

and then to the rest of the equation

$$-\frac{\hbar^2}{2}\left\{\frac{\left(\frac{1}{m_{i+1}}\frac{\psi_{i+2}-\psi_i}{2\delta z}\right)-\left(\frac{1}{m_{i-1}}\frac{\psi_{i-2}-\psi_i}{2\delta z}\right)}{2\delta z}\right\}+V_i\psi_i=E\psi_i \tag{2.6}$$

where ψ, m and V have been spatially discretised onto a grid of points each separated by the distance δz. Rearranging and removing a factor of two from δz gives

$$-\frac{\hbar^2}{2(\delta z)^2}\left\{\frac{\psi_{i+1}-\psi_i}{m_{i+\frac{1}{2}}}-\frac{\psi_i-\psi_{i-1}}{m_{i-\frac{1}{2}}}\right\}+V_i\psi_i=E\psi_i. \tag{2.7}$$

The $m\pm\frac{1}{2}$ points, which lie in between grid points, can be found by taking the mean of the two adjacent points since the FDA assumes a linear variation of variables between grid points.

While there are several methods for solving the discretised SE such as the shooting method [41] or a bi-section root finding approach [46], a matrix-method solution is used within this project as it presents several advantages:

- Reliability. Root finding approaches suffer from problems when the separation in energy between two states is smaller than resolution of the energy step used in the energy search. Since these algorithms rely on the change in sign of the wave function at one side of the simulation domain when they step through an energy solutions, they often miss pairs of states making them unreliable. Since diagonalisation of the matrix within the eigen value problem (EVP) will give all the eigenvalues of interest (i.e. those below the top of the conduction band edge), so long as the spatial step, δz, is small enough it is the most reliable method of solving the finite-difference Schrödinger equation (FDSE).

- Accuracy. Digitalisation of the EVP matrix will give the eigenvalues down to machine precision, again assuming that the spatial step is small. Root finding approaches rely on multiple iterations to home in on a particular solution therefore making an arbitrary precision sometimes impractical to achieve.

- Speed. Following on from the first two points; for a root finding approach to be both accurate and reliable it requires many iterations with a very small energy step, therefore making it slow. This implies that for root-finding approaches one of the three points must be sacrificed, however with a matrix method the solution may be all three since it will always take the same number of computational steps to achieve a reliable solution with arbitrary accuracy[3].

In order use matrix methods to solve (2.7), it must first be rearranged into a matrix equation of the form

$$\mathbf{H}|\psi\rangle = E|\psi\rangle, \tag{2.8}$$

where Dirac notation has been used to represent ψ as column vectors of spatial points and $\mathbf{H}$ is the spatially varying matrix representation of the Hamiltonian operator. Rearranging (2.7) to gather the coefficients of ψ gives

$$-\frac{\hbar^2}{2(\delta z)^2}\left\{\frac{\psi_{i+1}m_{i-\frac{1}{2}} - \psi_i(m_{i+\frac{1}{2}} + m_{i-\frac{1}{2}}) + \psi_{i-1}m_{i+\frac{1}{2}}}{m_{i+\frac{1}{2}}m_{i-\frac{1}{2}}}\right\} + V_i\psi_i = E\psi_i, \tag{2.9}$$

$$-\frac{\hbar^2}{2(\delta z)^2 m_{i+\frac{1}{2}}}\psi_{i+1} + \left\{\frac{\hbar^2}{2(\delta z)^2}\frac{m_{i+\frac{1}{2}} + m_{i-\frac{1}{2}}}{m_{i+\frac{1}{2}}m_{i-\frac{1}{2}}} + V_i\right\}\psi_i - \frac{\hbar^2}{2(\delta z)^2 m_{i-\frac{1}{2}}}\psi_{i-1} = E\psi_i, \tag{2.10}$$

a set of N simultaneous equations, with $i = 1, \ldots, N$ corresponding to N discrete spatial grid point within the simulation domain, which may be arranged into an $N \times N$ matrix, with each row being a different equation.

Schrödinger equation boundary conditions

Equation (2.10) does not, at present, represent a complete set of simultaneous equations since at the boundaries, i.e. $i = 1$ and $i = N$, the points ϕ_0 and ϕ_{N+1} are currently undefined. In order to complete the system of equations, boundary conditions must be imposed by defining the values of these two ϕ-points. Since semiconductor heterostructures are made up of recurring periodic structures, one would assume that a cyclic, or Neumann boundary condition would be the most suitable choice. However, because the devices simulated within this thesis are driven by an external electric field, the entire bandstructure is effectively *tilted* such that the proceeding period is at a lower relative potential to the current period. This means that the bandstructure is no longer truly periodic and cyclic boundaries may not be used.

[3] As shown later in section 2.1.3, when solving the SE with nonparabolic dispersion, matrix methods do become relatively slow. Unfortunately, this is the sacrifice one makes for arbitrary accuracy and reliability and is discussed in more detail within that section

The other option is to remove any influence of *the outside world* from the system under consideration, effectively placing an infinite distance from any other source of electrical potential (since the quantum well is made of an electrical potential). Practically, it would be impossible to implement this exactly since it would require an infinite simulation domain, however an approximate implementation of this boundary condition may be realised if two things are considered. Firstly, that by placing the quantum well system within an infinite domain, it is ensured that all wavefunctions are wholly confined within the quantum system since the wavefunctions decay inside the potential barriers and the barrier extend to $\pm$ infinity, therefore making ϕ at $\pm\infty = 0$. Secondly, since the wavefunctions decay quickly within the barriers, the point at which $\phi \approx 0$ will be some finite distance from the edge of the quantum well system. Therefore an approximate implementation may be achieved by fixing the value of the wavefunctions to zero at the boundaries of the simulation domain and ensuring that the barriers at the edge of the quantum well system are thick enough for the wavefunctions to decay to approximately zero. (This barrier *padding* is the operating condition for this type of boundary as discussed later.) These *hardwall*, or Dirichlet boundaries are implemented by setting the points at the boundary, i.e. ϕ_0 and ϕ_{N+1}, to zero thereby giving a complete set of equations that may be solved. (This approximation is equivalent to placing the quantum well system between two infinite potential barriers, since a wavefunction cannot penetrate into an infinite potential barrier so is zero at the boundary, which has implications for the quantum states which are found above the quantum well system as discussed later.)

Now that there is a complete system of equations to solve, (2.10) may be arranged into a matrix equation, with the matrix, $\mathbf{H}$, being tridiagonal and of the form

$$\mathsf{H} = \begin{pmatrix} b_1 & c_1 & 0 & \dots & 0 \\ a_2 & b_2 & c_2 & \dots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \dots & a_{N-1} & b_{N-1} & c_{N-1} \\ 0 & \dots & 0 & a_N & b_N \end{pmatrix}$$

where

$$\begin{aligned} a_i &= -\frac{\hbar^2}{2(\delta z)^2 m_{i-\frac{1}{2}}} \\ b_i &= \frac{\hbar^2 (m_{i+\frac{1}{2}} + m_{i-\frac{1}{2}})}{2(\delta z)^2 m_{i+\frac{1}{2}} m_{i-\frac{1}{2}}} + V_i \\ c_i &= -\frac{\hbar^2}{2(\delta z)^2 m_{i+\frac{1}{2}}} \end{aligned} \tag{2.11}$$

since $a_i = c_{i-1}$, $\mathbf{H}$ is also symmetric and therefore only a or c needs to be computed before solving. Solutions may be calculated by sending the constructed matrix to a linear algebra routine such as those provided by LAPACK [47], which will diagonalise the matrix and return the eigenvalues and corresponding values.

Model validation

As the solution of any matrix eigen-value problem will result in the same number of eigenvalues and eigenvectors as there are rows within the matrix (N equations will give N solutions), the set of solutions must be scrutinised in order to determine which are the solutions of interest and which are simply produced through the computational method used. In the case of the Schrödinger equation, the solutions of interest will be the *physically realistic* solutions where as the rest of the solutions will manifest themselves as higher-energy solutions resulting from the Dirichlet boundaries acting as infinite potential barriers. The bottom five solutions for a single quantum well quantum well are shown in figure 2.2, where it can be clearly seen that the solutions above the top of the quantum well are simply artifacts of the simulation. While these are real solutions to the Schrödinger equation with the imposed boundary conditions, they are samples of the *continuum* of states above the quantum well which are not of interest and can be discarded. Therefore, if the range of energy solutions is limited to be within the bounds of the quantum well system being examined then the expected solutions are obtained.

As discussed previously, a sufficient amount of barrier *padding* must be added to each side of the quantum well system in order to stop the hardwall boundaries affecting the states found. Figure 2.3 shows the effect of of varying the width of the padding at the edges of the same 10 nm quantum well in the $GaAs/Al_{0.15}Ga_{0.85}As$ material system. This shows that a significant amount of padding needs to be added (>5 nm) in order to reduce the effect of the boundaries. However, since the diagonalisation of a tridiagonal matrix requires very little computational effort it is better to be over-cautious with the amount of padding and therefore 10 nm is used.

Table 2.1 compares the energy solutions found for a 10 nm quantum well in the $GaAs/Al_{0.15}Ga_{0.85}As$ material system with an analytical solution, showing that the method agrees well. Figure 2.4 shows the wavefunctions calculated via the matrix method which are of the correct shape.

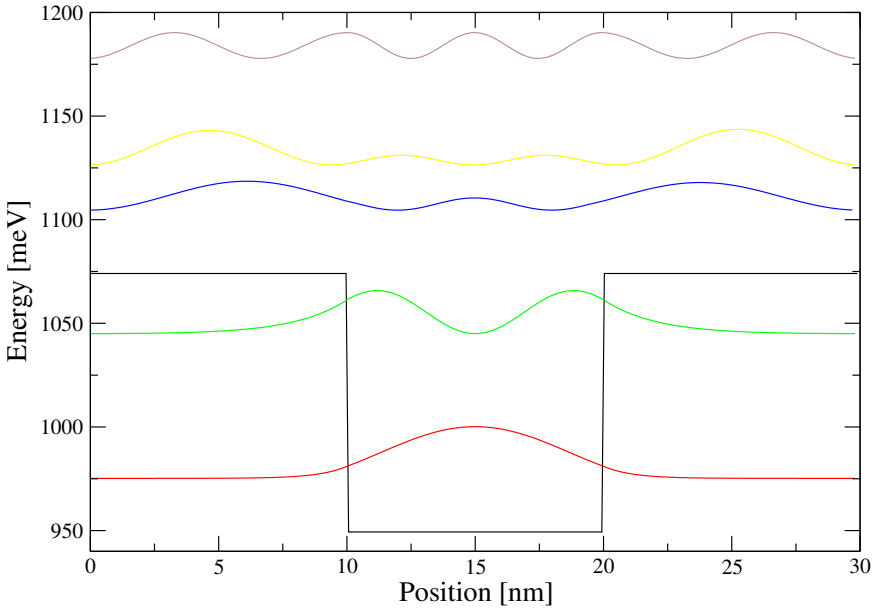

Figure 2.2: Wavefunction diagram of the lowest (in energy) five matrix solutions to the time-independent Schrödinger equation with parabolic effective mass within a 10 nm quantum well in the GaAs/$Al_{0.15}Ga_{0.85}As$ material system with 10 nm barriers with $\delta z = 0.1$ nm.

State	Analytical solution (meV)	Matrix solution (meV)	% error
1	975.07	975.19	0.46
2	1044.41	1045.03	0.66

Table 2.1: Comparison of matrix and analytical solutions to the time-independent Schrödinger equation with parabolic effective mass within a 10 nm quantum well in the GaAs/$Al_{0.15}Ga_{0.85}As$ material system. % error calculated from the bottom of the conduction band edge (949.33 meV) and taking the analytical solution to be correct.

2.1.3 Nonparabolic dispersion

Limitations of the effective mass approximation

As discussed before, the EMA is only valid for carriers at relatively low energies. This is not a problem for most semiconductor heterostructures because the band discontinuity between materials is only a few hundred milli-electron volts, meaning that the states of interest (i.e. confined states) cannot exist above this energy. There are some heterostructures however, such as InP based systems, where the band discontinuity can be of the order of one electron

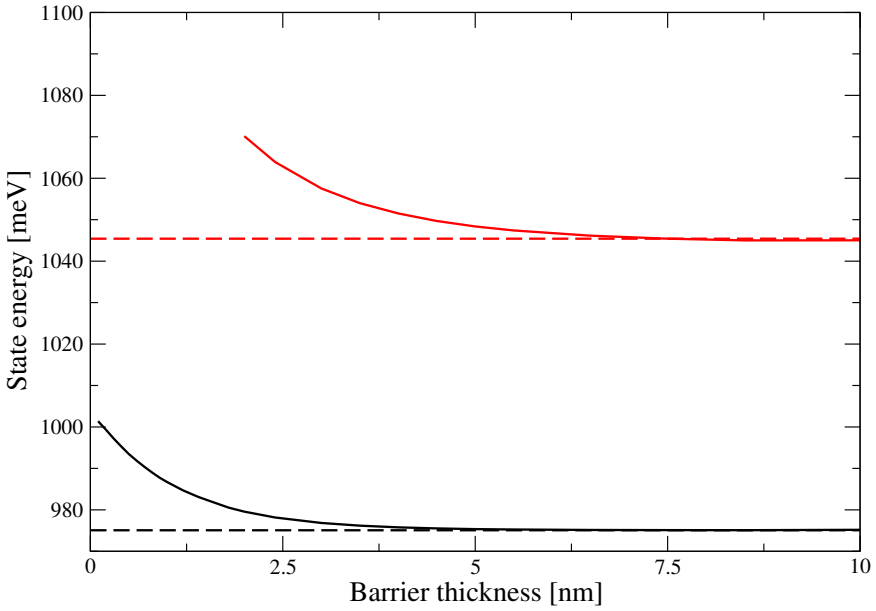

Figure 2.3: Effect of varying the padding barrier thickness on the energies of the states found within a 10 nm quantum well in the GaAs/$Al_{0.15}Ga_{0.85}As$ material system.

volt. For these systems the EMA goes far beyond its range of validity and does not produce reliable results. However, instead of moving to more complicated bandstructure simulation techniques, it is possible to improve the EMA such that it can be used to simulate much more energetic carriers within these types of systems.

In order to improve the EMA, consider why it is failing as carrier energies increase; the parabolic description of the dispersion relation is no longer close enough to the real dispersion relation to give an accurate representation of what is going on. It is possible to account for this deviation in real and parabolic dispersion relations by including an energy dependent term known as the *nonparabolicity*, as shown in Fig. 2.5.

In order to include this nonparabolicity within the effective mass approximation, the effective mass must become energy dependent, and take the form described by Nelson *et al.* [48]:

$$\frac{1}{m^*(E,z)} = \frac{1}{m^*(z)} \frac{1}{\{1+\alpha(z)[E-V(z)]\}} \tag{2.12}$$

where α is the nonparabolicity, an empirical fitting parameter which is also material dependent (hence the z dependence), E is the energy and V is band edge potential profile.

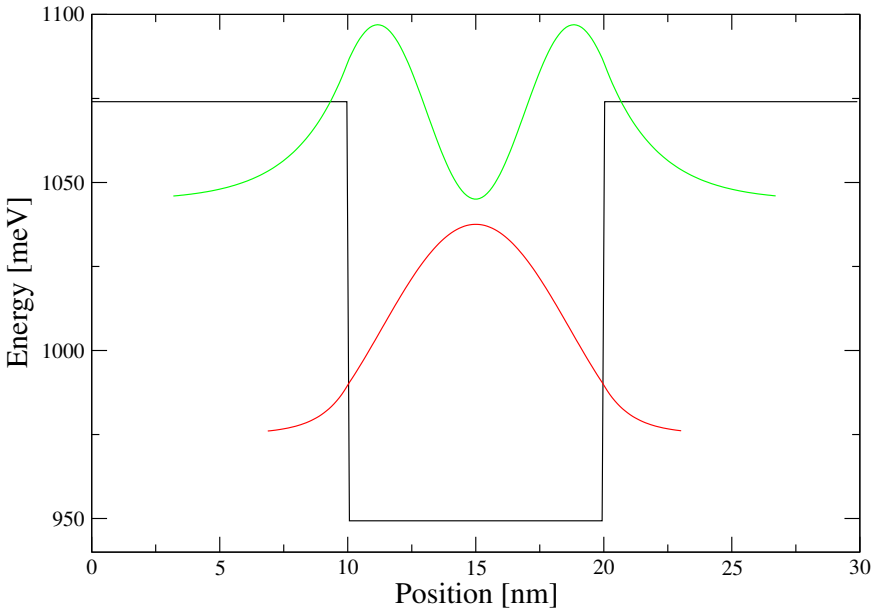

Figure 2.4: Wavefunction diagram of the matrix solutions to the time-independent Schrödinger equation with parabolic effective mass within a 10 nm quantum well in the GaAs/$Al_{0.15}Ga_{0.85}As$ material system.

Eigenvalue problem linearisation

Deriving a numerical solution for the nonparabolic SE, as in section 2.1.2, yields a cubic EVP because of the energy dependence of the nonparabolic effective mass. Therefore the problem cannot be solved directly be an eigenvalue solver and some form of root finding algorithm must be employed in order to to *search* for solutions. However, if the cubic EVP is linearised after it is derived, then a direct method of solution can still be used.

The derivation of the cubic EVP begins by inserting the nonparabolic effective mass into the SE:

$$\frac{-\hbar^2}{2}\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{m^*(z)\{1+\alpha(z)[E-V(z)]\}}\frac{\mathrm{d}}{\mathrm{d}z}\psi(z)\right)+V(z)\psi(z)=E\psi(z) \tag{2.13}$$

Discretising equation (2.13) gives:

$$\frac{-\hbar^2}{2\delta z^2}\left\{\frac{\psi_{i+1}-\psi_i}{m_{i+\frac{1}{2}}\left[1+\alpha_{i+\frac{1}{2}}\left(E-V_{i+\frac{1}{2}}\right)\right]}-\frac{\psi_i-\psi_{i-1}}{m_{i-\frac{1}{2}}\left[1+\alpha_{i-\frac{1}{2}}\left(E-V_{i-\frac{1}{2}}\right)\right]}\right\}+V_i\psi_i=E\psi_i \tag{2.14}$$

In order to rearrange this problem into an EVP that is solvable, we must gather all the terms of E through the equation. This requires the denominators of equation (2.13) to be multiplied

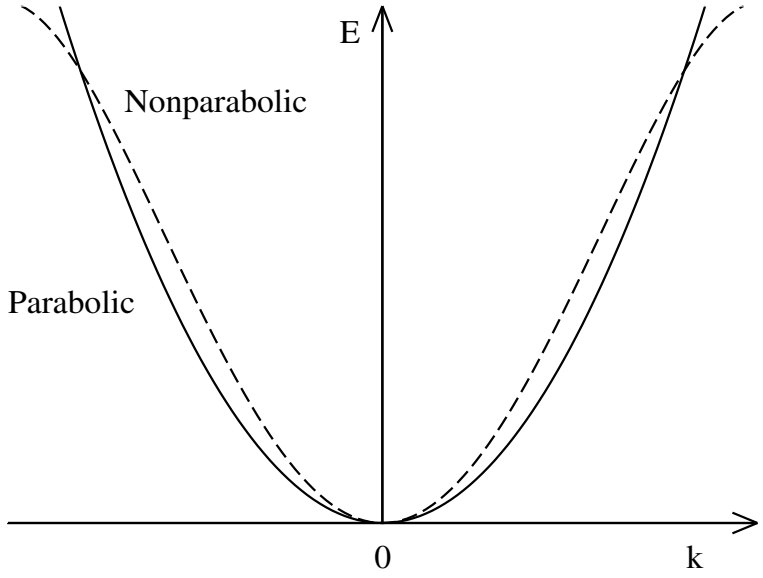

Figure 2.5: The parabolic and nonparabolic dispersion relations. Whilst the parabolic dispersion is close to the more physically-realistic nonparabolic dispersion, at higher wavevectors the nonparabolic dispersion *flatterns out* to give more realistic state energies for energies high above the conduction band.

through the entire equation, giving:

$$\begin{aligned}\frac{-\hbar^2}{2\delta z^2}\left\{m_{i-\frac{1}{2}}\left[1+\alpha_{i-\frac{1}{2}}\left(E-V_{i-\frac{1}{2}}\right)\right](\psi_{i+1}-\psi_i)-m_{i+\frac{1}{2}}\left[1+\alpha_{i+\frac{1}{2}}\left(E-V_{i+\frac{1}{2}}\right)\right](\psi_i-\psi_{i-1})\right\}\\ +m_{i+\frac{1}{2}}\left[1+\alpha_{i+\frac{1}{2}}\left(E-V_{i+\frac{1}{2}}\right)\right]m_{i-\frac{1}{2}}\left[1+\alpha_{i-\frac{1}{2}}\left(E-V_{i-\frac{1}{2}}\right)\right]V_i\psi_i\\ =m_{i+\frac{1}{2}}\left[1+\alpha_{i+\frac{1}{2}}\left(E-V_{i+\frac{1}{2}}\right)\right]m_{i-\frac{1}{2}}\left[1+\alpha_{i-\frac{1}{2}}\left(E-V_{i-\frac{1}{2}}\right)\right]E\psi_i\end{aligned} \tag{2.15}$$

Expanding all the $m_{i+\frac{1}{2}}(1+\alpha_{i+\frac{1}{2}}(E-V_{i+\frac{1}{2}}))m_{i-\frac{1}{2}}(1+\alpha_{i-\frac{1}{2}}(E+V_{i-\frac{1}{2}}))$ terms, then gathering

all the coefficients of powers of E and arranging them by coefficients of ψ gives:

$$
\begin{aligned}
&m_{i+\frac{1}{2}} m_{i-\frac{1}{2}} \alpha_{i+\frac{1}{2}} \alpha_{i-\frac{1}{2}} E^3 \psi_i \\
&+ \left\{ \alpha_{i-\frac{1}{2}} + \alpha_{i+\frac{1}{2}} - \alpha_{i+\frac{1}{2}} \alpha_{i-\frac{1}{2}} (V_{i+\frac{1}{2}} + V_i + V_{i-\frac{1}{2}}) \right\} m_{i+\frac{1}{2}} m_{i-\frac{1}{2}} E^2 \psi_i \\
&+ \Bigg\{ \frac{\hbar^2}{2\delta z^2} m_{i-\frac{1}{2}} \alpha_{i-\frac{1}{2}} \psi_{i+1} \\
&\quad + \Big[\frac{\hbar^2}{2\delta z^2} \big(-m_{i-\frac{1}{2}} \alpha_{i-\frac{1}{2}} - m_{i+\frac{1}{2}} \alpha_{i+\frac{1}{2}} \big) + \big(1 - \alpha_{i-\frac{1}{2}} (V_{i-\frac{1}{2}} + V_i) - \alpha_{i+\frac{1}{2}} (V_{i+\frac{1}{2}} + V_i) \\
&\qquad + \alpha_{i+\frac{1}{2}} \alpha_{i-\frac{1}{2}} (V_{i+\frac{1}{2}} V_i + V_{i-\frac{1}{2}} V_i + V_{i+\frac{1}{2}} V_{i-\frac{1}{2}}) \big) m_{i+\frac{1}{2}} m_{i-\frac{1}{2}} \Big] \psi_i \\
&\quad + \frac{\hbar^2}{2\delta z^2} m_{i+\frac{1}{2}} \alpha_{i+\frac{1}{2}} \psi_{i-1} \Bigg\} E \\
&+ \Bigg\{ \frac{\hbar^2}{2\delta z^2} m_{i-\frac{1}{2}} (1 - \alpha_{i-\frac{1}{2}} V_{i-\frac{1}{2}}) \psi_{i+1} \\
&\quad + \Big[\frac{\hbar^2}{2\delta z^2} \big(-m_{i-\frac{1}{2}} (1 - \alpha_{i-\frac{1}{2}} V_{i-\frac{1}{2}}) - m_{i+\frac{1}{2}} (1 - \alpha_{i+\frac{1}{2}} V_{i+\frac{1}{2}}) \big) \\
&\qquad - \Big(1 - \alpha_{i-\frac{1}{2}} V_{i-\frac{1}{2}} - \alpha_{i+\frac{1}{2}} V_{i+\frac{1}{2}} + \alpha_{i+\frac{1}{2}} \alpha_{i-\frac{1}{2}} V_{i+\frac{1}{2}} V_{i-\frac{1}{2}} \Big) m_{i+\frac{1}{2}} m_{i-\frac{1}{2}} V_i \Big] \psi_i \\
&\quad + \frac{\hbar^2}{2\delta z^2} m_{i+\frac{1}{2}} (1 - \alpha_{i+\frac{1}{2}} V_{i+\frac{1}{2}}) \psi_{i-1} \Bigg\} = 0
\end{aligned}
\tag{2.16}
$$

which is a cubic EVP of the form:

$$
\left(E^3 \mathsf{A_4} + E^2 \mathsf{A_3} + E \mathsf{A_2} + \mathsf{A_1} \right) \psi = 0 \tag{2.17}
$$

Applying the same hardwall boundary conditions derived in 2.1.2, the matrices $\mathsf{A_4}$, $\mathsf{A_3}$, $\mathsf{A_2}$ and $\mathsf{A_1}$ are all either diagonal or tridiagonal symmetric with the forms:

$$
\mathsf{A_1} = \begin{pmatrix} b_1 & c_1 & 0 & \dots & 0 \\ a_2 & b_2 & c_2 & \dots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \dots & a_{N-1} & b_{N-1} & c_{N-1} \\ 0 & \dots & 0 & a_N & b_N \end{pmatrix}, \quad \mathsf{A_2} = \begin{pmatrix} e_1 & f_1 & 0 & \dots & 0 \\ d_2 & e_2 & f_2 & \dots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \dots & d_{N-1} & e_{N-1} & f_{N-1} \\ 0 & \dots & 0 & d_N & e_N \end{pmatrix},
$$

$$
\mathsf{A_3} = \begin{pmatrix} g_1 & 0 & \dots & 0 \\ 0 & g_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & g_N \end{pmatrix} \quad \text{and} \quad \mathsf{A_4} = \begin{pmatrix} h_1 & 0 & \dots & 0 \\ 0 & h_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & h_N \end{pmatrix}, \tag{2.18}
$$

with matrix elements given by:

$$
\begin{aligned}
a_i &= \frac{\hbar^2}{2(\delta z)^2} m_{i+\frac{1}{2}} \left(1 - \alpha_{i+\frac{1}{2}} V_{i+\frac{1}{2}}\right) \\
b_i &= \frac{\hbar^2}{2(\delta z)^2} \left[-m_{i-\frac{1}{2}} \left(1 - \alpha_{i-\frac{1}{2}} V_{i-\frac{1}{2}}\right) - m_{i+\frac{1}{2}} \left(1 - \alpha_{i+\frac{1}{2}} V_{i+\frac{1}{2}}\right)\right] \\
&\quad - \left(1 - \alpha_{i-\frac{1}{2}} V_{i-\frac{1}{2}} - \alpha_{i+\frac{1}{2}} V_{i+\frac{1}{2}} + \alpha_{i+\frac{1}{2}} \alpha_{i-\frac{1}{2}} V_{i+\frac{1}{2}} V_{i-\frac{1}{2}}\right) m_{i+\frac{1}{2}} m_{i-\frac{1}{2}} V_i \\
c_i &= \frac{\hbar^2}{2(\delta z)^2} m_{i-\frac{1}{2}} \left(1 - \alpha_{i-\frac{1}{2}} V_{i-\frac{1}{2}}\right) \\
d_i &= \frac{\hbar^2}{2(\delta z)^2} m_{i+\frac{1}{2}} \alpha_{i+\frac{1}{2}} \\
e_i &= \frac{\hbar^2}{2(\delta z)^2} \left(-m_{i-\frac{1}{2}} \alpha_{i-\frac{1}{2}} - m_{i+\frac{1}{2}} \alpha_{i+\frac{1}{2}}\right) \\
&\quad + \left[1 - \alpha_{i-\frac{1}{2}} \left(V_{i-\frac{1}{2}} + V_i\right) - \alpha_{i+\frac{1}{2}} \left(V_{i+\frac{1}{2}} + V_i\right)\right. \\
&\quad \left. + \alpha_{i+\frac{1}{2}} \alpha_{i-\frac{1}{2}} \left(V_{i+\frac{1}{2}} V_i + V_{i-\frac{1}{2}} V_i + V_{i+\frac{1}{2}} V_{i-\frac{1}{2}}\right)\right] m_{i+\frac{1}{2}} m_{i-\frac{1}{2}} \\
f_i &= \frac{\hbar^2}{2(\delta z)^2} m_{i-\frac{1}{2}} \alpha_{i-\frac{1}{2}} \\
g_i &= \left[\alpha_{i-\frac{1}{2}} + \alpha_{i+\frac{1}{2}} - \alpha_{i+\frac{1}{2}} \alpha_{i-\frac{1}{2}} \left(V_{i+\frac{1}{2}} + V_i + V_{i-\frac{1}{2}}\right)\right] m_{i+\frac{1}{2}} m_{i-\frac{1}{2}} \\
h_i &= m_{i+\frac{1}{2}} m_{i-\frac{1}{2}} \alpha_{i+\frac{1}{2}} \alpha_{i-\frac{1}{2}}
\end{aligned}
\tag{2.19}
$$

This cubic EVP can be recast to give a linear, but *generalised* EVP of the form:

$$
\begin{pmatrix} 0 & \mathsf{I_N} & 0 \\ 0 & 0 & \mathsf{I_N} \\ \mathsf{A_1} & \mathsf{A_2} & \mathsf{A_3} \end{pmatrix} \begin{pmatrix} \psi \\ E\psi \\ E^2\psi \end{pmatrix} = E \begin{pmatrix} \mathsf{I_N} & 0 & 0 \\ 0 & \mathsf{I_N} & 0 \\ 0 & 0 & -\mathsf{A_4} \end{pmatrix} \begin{pmatrix} \psi \\ E\psi \\ E^2\psi \end{pmatrix} \tag{2.20}
$$

where $\mathsf{I_N}$ is the identity matrix with a leading dimension N. As the matrix on the right hand side of (2.20) is diagonal, its inverse is also diagonal, and one can transform (2.20) into a normal EVP by pre-multiplying by the right hand matrix:

$$
\begin{pmatrix} 0 & \mathsf{I_N} & 0 \\ 0 & 0 & \mathsf{I_N} \\ -\mathsf{A_4}^{-1}\mathsf{A_1} & -\mathsf{A_4}^{-1}\mathsf{A_2} & -\mathsf{A_4}^{-1}\mathsf{A_3} \end{pmatrix} \begin{pmatrix} \psi \\ E\psi \\ E^2\psi \end{pmatrix} = E \begin{pmatrix} \psi \\ E\psi \\ E^2\psi \end{pmatrix} \tag{2.21}
$$

This pre-multiplication by $\mathsf{A_4}^{-1}$ is the equivalent of dividing through all the terms in (2.19)

by the h term. This modifies the terms in (2.19) to

$$
\begin{aligned}
a_i &= -\frac{\hbar^2(1-\alpha_{i+\frac{1}{2}}V_{i+\frac{1}{2}})}{2(\delta z)^2 m_{i-\frac{1}{2}}\alpha_{i+\frac{1}{2}}\alpha_{i-\frac{1}{2}}} \\
b_i &= \frac{\hbar^2}{2(\delta z)^2\alpha_{i+\frac{1}{2}}\alpha_{i-\frac{1}{2}}}\left(\frac{1-\alpha_{i-\frac{1}{2}}V_{i-\frac{1}{2}}}{m_{i+\frac{1}{2}}}+\frac{1-\alpha_{i+\frac{1}{2}}V_{i+\frac{1}{2}}}{m_{i-\frac{1}{2}}}\right) \\
&\quad + \frac{V_i}{\alpha_{i+\frac{1}{2}}\alpha_{i-\frac{1}{2}}}\left(1-\alpha_{i-\frac{1}{2}}V_{i-\frac{1}{2}}-\alpha_{i+\frac{1}{2}}V_{i+\frac{1}{2}}+\alpha_{i+\frac{1}{2}}\alpha_{i-\frac{1}{2}}V_{i+\frac{1}{2}}V_{i-\frac{1}{2}}\right) \\
c_i &= -\frac{\hbar^2(1-\alpha_{i-\frac{1}{2}}V_{i-\frac{1}{2}})}{2(\delta z)^2 m_{i+\frac{1}{2}}\alpha_{i+\frac{1}{2}}\alpha_{i-\frac{1}{2}}} \\
d_i &= -\frac{\hbar^2}{2(\delta z)^2 m_{i-\frac{1}{2}}\alpha_{i-\frac{1}{2}}} \\
e_i &= \frac{\hbar^2}{2(\delta z)^2}\left(\frac{1}{m_{i+\frac{1}{2}}\alpha_{i+\frac{1}{2}}}+\frac{1}{m_{i-\frac{1}{2}}\alpha_{i-\frac{1}{2}}}\right) \\
&\quad - \frac{1}{\alpha_{i+\frac{1}{2}}\alpha_{i-\frac{1}{2}}}\left[1-\alpha_{i-\frac{1}{2}}\left(V_{i-\frac{1}{2}}+V_i\right)-\alpha_{i+\frac{1}{2}}\left(V_{i+\frac{1}{2}}+V_i\right)\right. \\
&\quad \left.+\alpha_{i+\frac{1}{2}}\alpha_{i-\frac{1}{2}}\left(V_{i+\frac{1}{2}}V_i+V_{i-\frac{1}{2}}V_i+V_{i+\frac{1}{2}}V_{i-\frac{1}{2}}\right)\right] \\
f_i &= -\frac{\hbar^2}{2(\delta z)^2 m_{i+\frac{1}{2}}\alpha_{i+\frac{1}{2}}} \\
g_i &= -\frac{1}{\alpha_{i+\frac{1}{2}}}-\frac{1}{\alpha_{i-\frac{1}{2}}}+\left(V_{i+\frac{1}{2}}+V_i+V_{i-\frac{1}{2}}\right)
\end{aligned}
\tag{2.22}
$$

with the h term no longer being used, such that the terms sit in the matrices A_1, A_2 and A_3 in the same way shown in (2.18). Note that A_2 is actually a tridiagonal symmetric matrix, however all terms have been specified here since any practice implementation of constructing this matrix will scan across the simulation, calculating the $i \pm \frac{1}{2}$ values at each point, and therefore it is more convenient to define the whole matrix.

The cubic EVP has now been recast into a sparse, banded, linear EVP, with the matrix here being three times larger than those of the cubic problem. This EVP can be easily solved by any appropriate EVP solver, such as one from the LAPACK library [47]. It should be noted that setting α to zero in equations (2.22) reduces to the nonparabolic formulation to the parabolic formulation derived in section 2.1.2. This gives a good indication that the derivation presented above is correct, as one would expect that setting α to zero, and therefore removing the nonparabolic component from the EMA, would give exactly the same solution as just solving the parabolic SE.

Model validation

In order to ensure the validity of proposed method, its results are first compared to an analytical solution for a single quantum well which includes nonparabolicity. This is done within the $In_{0.53}Ga_{0.47}As/AlAs_{0.56}Sb_{0.44}$ material system because of its very high conduction band offset of 1.6 eV which accentuates the effect of nonparabolicity. Table 2.2 shows a comparison of the energy levels found in a 10 nm quantum well by the presented method, a numerical solution of the transcendental equation with an energy dependent [49], i.e. nonparabolic, effective mass, and the parabolic method presented in 2.1.2, to show how prominent an effect nonparabolicity is high above the conduction band, the parabolic method from 2.1.2. Clearly, the parabolic method overestimates the energy solutions (except for the ground state where, due to the nonparabolic dispersion giving a marginally higher energy at low k-vectors as in 2.5, the energy is underestimated slightly), and the error increases for solutions that are higher above the conduction band edge. The overestimation of the energy solutions from the parabolic method is so prominent towards the top of the quantum well that the top three solutions are not found. Conversely, the nonparabolic method shows excellent agreement with the analytical solution with the absolute difference being only a few milli-electron volts for all solutions. Figure 2.6 shows the solutions found by the method confirming that the wavefunctions are of the expected form.

In order to rigorously test the shape of the wavefunctions produced by the method, the orthogonality of two wavefunctions from a single quantum well are tested. However, because the nonparabolicity inherently introduces nonorthogonality, the nonparabolicity parameter, α, was set set to zero in order to remove nonparabolic. Therefore, if the shape of the wavefunctions found by presented method is correct, then with $\alpha = 0$ the states within a single quantum well strucutre should be exactly orthogonal (i.e. $\langle i|j\rangle = 0$). The orthogonality of the solutions found by the presented method, with $\alpha = 0$, is compared to that of the widely used parabolic shooting method [41], which utilises a bisection search root-finding algorithm to find solutions. The comparison is done for a varying number of sampling points within a fixed size system, and therefore different spatial sampling steps δz, as well as different values of the bisection search accuracy limit which determines to what precision the solutions are found within the search algorithm. Figure 2.7 shows the orthogonality of the top two solutions found by the two methods, $\langle 4|5\rangle$, for the 20 nm single quantum well show in the inset of 2.7. Since the matrix method is a *direct* solution, it will always produce solutions that are orthogonal down to machine precision and therefore $\langle 4|5\rangle$ has a constant value of 4×10^{-17}.

State	Transcendental solution (meV)	Parabolic solution (meV)	Parabolic % error	Presented method (meV)	Presented method % error
1	829.28	823.36	9.41	830.18	1.43
2	983.64	998.50	6.84	985.68	0.94
3	1182.21	1300.61	28.47	1184.68	0.59
4	1400.86	1734.01	52.51	1403.29	0.38
5	1628.90	2272.80	74.66	1631.12	0.26
6	1860.78	-	-	1862.75	0.18
7	2092.03	-	-	2093.80	0.13
8	2313.78	-	-	2314.35	0.04

Table 2.2: Comparison of transcendental, parabolic and nonparabolic (i.e. the Presented method) solutions to the time-independent Schrödinger equation with parabolic effective mass within a 10 nm quantum well in the $In_{0.53}Ga_{0.47}As/AlAs_{0.56}Sb_{0.44}$ material system. % error calculated from the bottom of the conduction band edge (766.40 meV) and taking the analytical solution to be correct.

The shooting method on the other hand, produces solutions whose orthogonality depends very heavily on the number of sampling points used and reaches a minimum orthogonality, $\langle 4|5\rangle$, of 0.002. The vastly lower orthogonality of the solutions from the matrix method imply that the wavefunctions produced by this method have a much higher accuracy than those produced by the shooting method.

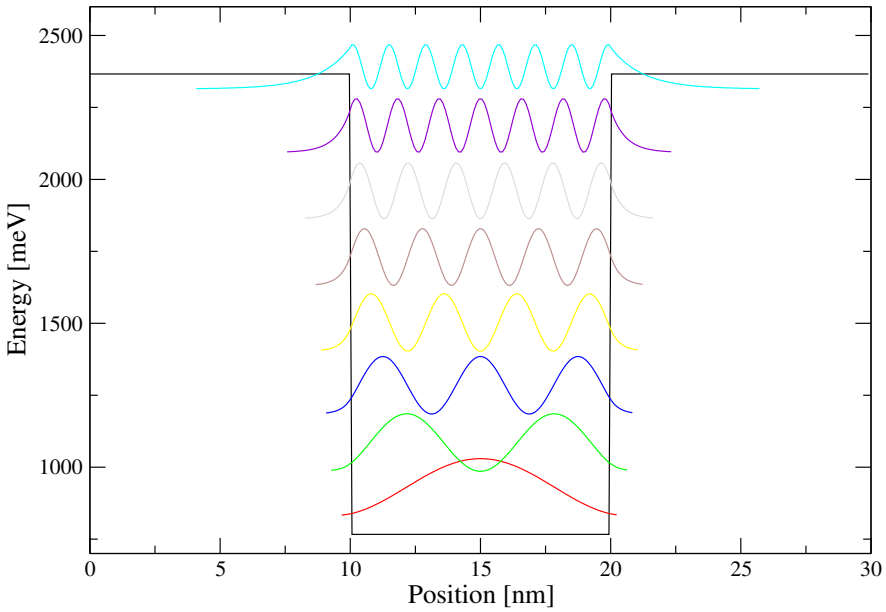

Figure 2.6: Wavefunction diagram of the nonparabolic solutions to the time-independent Schrödinger equation within a 10 nm quantum well in the $In_{0.53}Ga_{0.47}As/AlAs_{0.56}Sb_{0.44}$ material system. Wavefunction tails have been cutoff for probability densities below 1%

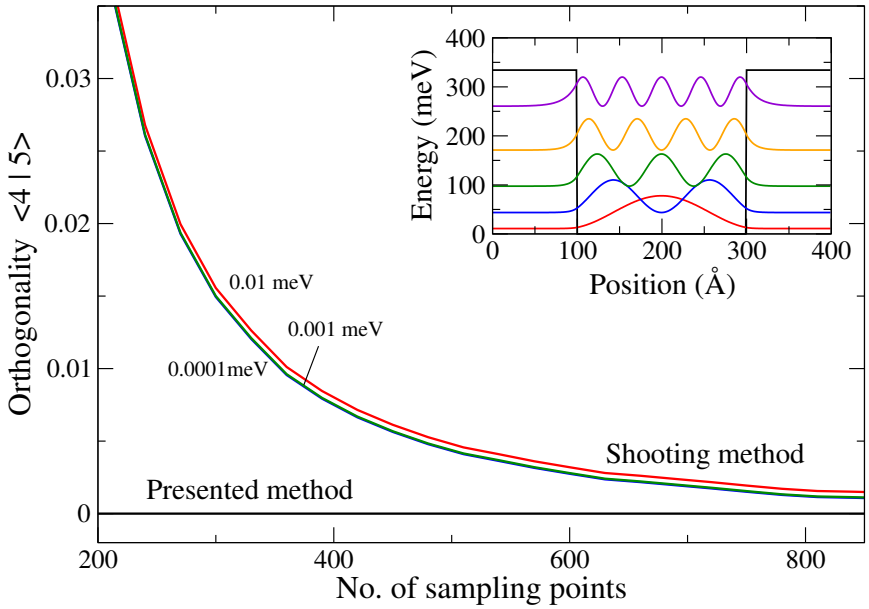

Figure 2.7: Orthogonality of wavefunctions, $\langle 4|5\rangle$, with respect to the number of sampling points for the presented method, with $\alpha = 0$, and the parabolic shooting method [41] for several values of bisection search accuracy. (Inset) The 20 nm single quantum well under consideration in the $GaAs/Al_{0.33}Ga_{0.77}As$ material system.

2.2 Poisson's equation in one dimension

Semiconductor heterostructures often make use of dopant atoms at specific points within the device, acting as either donors or acceptors, to increase the number electrons or holes respectively. Once these dopants donate or accept an electron not only do they become ionised, but they also change the amount of charge within the conduction or valence band. Furthermore, since electrons and holes are free to move through the heterostructure (via tunnelling in the direction of confinement), whereas the ionised dopants have a fixed position within the crystal lattice, there will no longer be charge neutrality, locally, along the heterostructure. This net charge will induce voltages locally along the heterostructure which will bend the bandstrucutre and, in tern, shift the positions of the electrons within the sturucture (or rather shift the expectiation positions of the wavefucntions). The movement of electrons will affect the charge distribution, inducing a different voltage which will further alter the shape of the bandstrucutre.

In one dimension, the spatial profile of an induced voltage, ϕ, is related to the charge distribution which induced it, ρ, via Poisson's equation,

$$\frac{\partial}{\partial z}\left(\varepsilon(z)\frac{\partial}{\partial z}\phi(z)\right) = -\rho \tag{2.23}$$

where z is the axis of the one extended dimension and ε is the spatially dependent effective permittivity.

Applying the finite difference approximation, to the inner-most differential first,

$$\frac{\partial}{\partial z}\left(\varepsilon(z)\frac{\phi_{i+1}-\phi_{i-1}}{2\delta z}\right) = -\rho, \tag{2.24}$$

and then to the rest of (2.23)

$$\frac{\varepsilon_{i+1}\left(\frac{\phi_{i+2}-\phi_i}{2\delta z}\right) - \varepsilon_{i-1}\left(\frac{\phi_i-\phi_{i-2}}{2\delta z}\right)}{2\delta z} = -\rho_i, \tag{2.25}$$

where δz is the spatial step separating the grid points $i = 1, \ldots, N,$. Removing the factor of 2 from δz, and rearranging to gather the coefficients of ϕ gives,

$$\frac{1}{\delta z}\left\{\varepsilon_{i+\frac{1}{2}}\phi_{i+1} - (\varepsilon_{i+\frac{1}{2}} + \varepsilon_{i-\frac{1}{2}})\phi_i + \varepsilon_{i-\frac{1}{2}}\phi_{i-1}\right\} = -\rho_i, \tag{2.26}$$

where the $i \pm \frac{1}{2}$ terms may be found by taking the average of the adjacent points.

By arranging ϕ and ρ into column vectors of the form,

$$\phi^{\mathsf{T}} = \{\phi_1, \phi_2, \ldots, \phi_N\},$$

$$\rho^{\mathsf{T}} = \{\rho_1, \rho_2, \ldots, \rho_N\},$$

(2.26) may be rewritten as a matrix equation of the form

$$\mathbf{A}\phi = -\rho, \tag{2.27}$$

where the exact form of the matrix $\mathbf{A}$ is determined by the boundary conditions imposed.

2.2.1 Boundary conditions

Hardwall boundary conditions

Hardwall, or Dirichlet boundary conditions force the value of ϕ to be a particular value (normally zero) at the edges of the simulation domain. Imposing this condition, i.e. $\phi_0 = \phi_{N+1} = 0$, we may write down a complete set of N equations to describe the potential resulting from some charge distribution, discretised onto a grid of N points:

$$\begin{aligned}
\frac{1}{\delta z}\left\{\varepsilon_{i+\frac{1}{2}}\phi_{i+1} - (\varepsilon_{i+\frac{1}{2}} + \varepsilon_{i-\frac{1}{2}})\phi_i\right\} &= -\rho_i, \qquad \text{for } i = 1; \\
\frac{1}{\delta z}\left\{\varepsilon_{i+\frac{1}{2}}\phi_{i+1} - (\varepsilon_{i+\frac{1}{2}} + \varepsilon_{i-\frac{1}{2}})\phi_i + \varepsilon_{i-\frac{1}{2}}\phi_{i-1}\right\} &= -\rho_i, \qquad \text{for } i = 2, 3, \ldots, N-1; \\
\frac{1}{\delta z}\left\{-(\varepsilon_{i+\frac{1}{2}} + \varepsilon_{i-\frac{1}{2}})\phi_i + \varepsilon_{i-\frac{1}{2}}\phi_{i-1}\right\} &= -\rho_i, \qquad \text{for } i = N;
\end{aligned} \tag{2.28}$$

where the effective permittivity may be assumed constant across the boundaries of the simulation domain for $\varepsilon_{i=\frac{1}{2}}$ and $\varepsilon_{i=N+\frac{1}{2}}$.

Rearranging into a matrix equation of the form (2.27), the matrix to solve, A takes the form

$$\mathsf{A} = \begin{pmatrix}
b_1 & a_1 & 0 & \ldots & 0 \\
a_1 & b_2 & a_2 & \ldots & \vdots \\
0 & \ddots & \ddots & \ddots & 0 \\
\vdots & \ldots & a_{N-2} & b_{N-1} & a_{N-1} \\
0 & \ldots & 0 & a_{N-1} & b_N
\end{pmatrix}, \tag{2.29}$$

where

$$\begin{aligned}
a_i &= \frac{1}{\delta z}\varepsilon_{i+\frac{1}{2}} \\
b_i &= -\frac{1}{\delta z}(\varepsilon_{i+\frac{1}{2}} + \varepsilon_{i-\frac{1}{2}})
\end{aligned} \tag{2.30}$$

Since A is a tridiagonal symmetric matrix it may be solved rapidly using the Thomas algorithm, which is implemented in many linear algebra software packages such as LAPACK [47].

In order to test the solution Poisson's equation, a 10 nm single quantum well in the $GaAs/Al_{0.15}Ga_{0.85}As$ material system is considered, with both of the barriers being n-type doped with a doping density of $1\times10^{18}\,\mathrm{cm}^{-3}$. As discussed previously, these dopant atoms will donate free electrons whose position will clearly be dependant upon the wavefunctions found

by the solution to the Schrödinger equation. However, for simplicity, a first approximation may be made which does not require the solution of the Schrödinger equation by assuming that the electrons will be evenly distributed throughout the heterostructure. Figure 2.8 shows the resulting Poisson potential, ϕ, as well as the total band-edge potential, V_{total}, which is the sum of ϕ and the conduction band potential, V_{CB}.

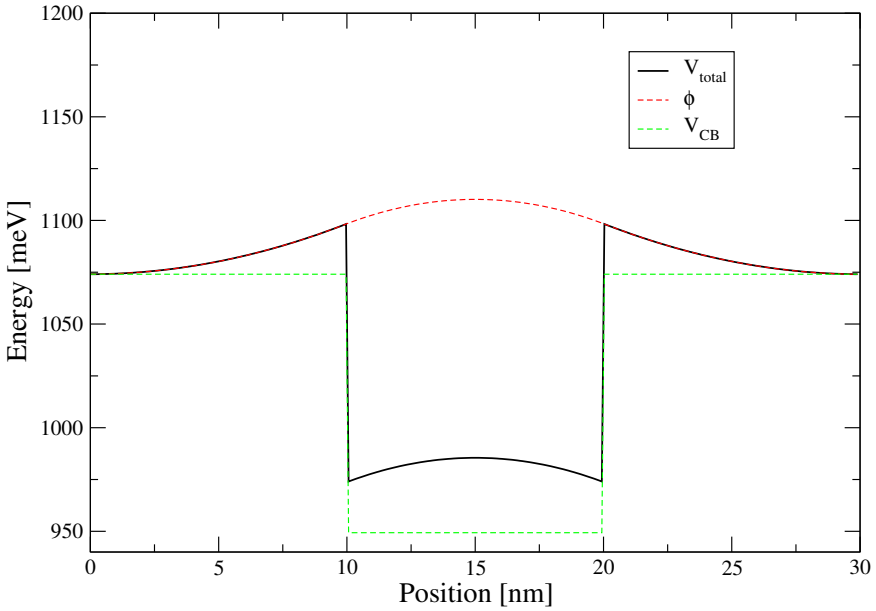

Figure 2.8: The Poisson potential, ϕ, for a 10 nm quantum well in the GaAs/$Al_{0.15}Ga_{0.85}$As material system with barriers which are n-type doped to $1 \times 10^{18}\,\text{cm}^{-3}$, assuming all donors are ionised and free-carriers are evenly distributed throughout the entire structure. The total band-edge profile, V_{total}, is also shown which is the sum of ϕ and the conduction band edge, V_{CB}. (The Poisson potential has been offset by the maximum value of the conduction band edge for clarity.)

As many semiconductor heterostructures are placed under some applied bias for normal operation, it is useful to define how the potential may be fixed at the boundaries using the Dirichlet boundary conditions. Furthermore, it is more accurate to apply a bias using these boundary conditions than to simply add a linearly increasing/decreasing potential to the conduction band edge, as the former method will account for the changes in the effective permeability throughout the structure (and is in fact equivalent to solving Poisson's equation followed by Laplace's equation and adding the results).

If the boundary points are fixed to some potential, i.e. $\phi_0 \neq 0,\ \phi_{N+1} \neq 0$, the equations

at $i = 1$ and $i = N$ become

$$\frac{1}{\delta z}\left\{\varepsilon_{\frac{1}{2}}\phi_2 - (\varepsilon_{1\frac{1}{2}} + \varepsilon_{\frac{1}{2}})\phi_1\varepsilon_{\frac{1}{2}}\phi_0\right\} = -\rho_1, \tag{2.31}$$

and

$$\frac{1}{\delta z}\left\{\varepsilon_{N+\frac{1}{2}}\phi_{N+1} - (\varepsilon_{N+\frac{1}{2}} + \varepsilon_{N-\frac{1}{2}})\phi_N + \varepsilon_{N-\frac{1}{2}}\phi_{N-1}\right\} = -\rho_N, \tag{2.32}$$

respectively. Since there is no space within the matrix, $\mathbf{A}$, for the ϕ_0 and ϕ_{N+1} terms, and as the column vector ϕ is still unknown, these terms must be moved over to right hand side when these equations are arranged within the matrix equation (2.27). This alters the form of the column vector, ρ, such that it becomes

$$\rho^{\mathsf{T}} = \{\rho_1 + \frac{1}{\delta z}\varepsilon_{\frac{1}{2}}\phi_0, \rho_2, \ldots, \rho_N + \frac{1}{\delta z}\varepsilon_{N+\frac{1}{2}}\phi_{N+1}\}, \tag{2.33}$$

where the points which lie outside the simulation domain, $\varepsilon_{\frac{1}{2}}$ and $\varepsilon_{N+\frac{1}{2}}$, can assumed to be identical to their adjacent points which lie inside the simulation domain. Figure 2.9 shows the same 10 nm quantum well as in 2.8 with an applied bias of 15 kVcm^{-1}.

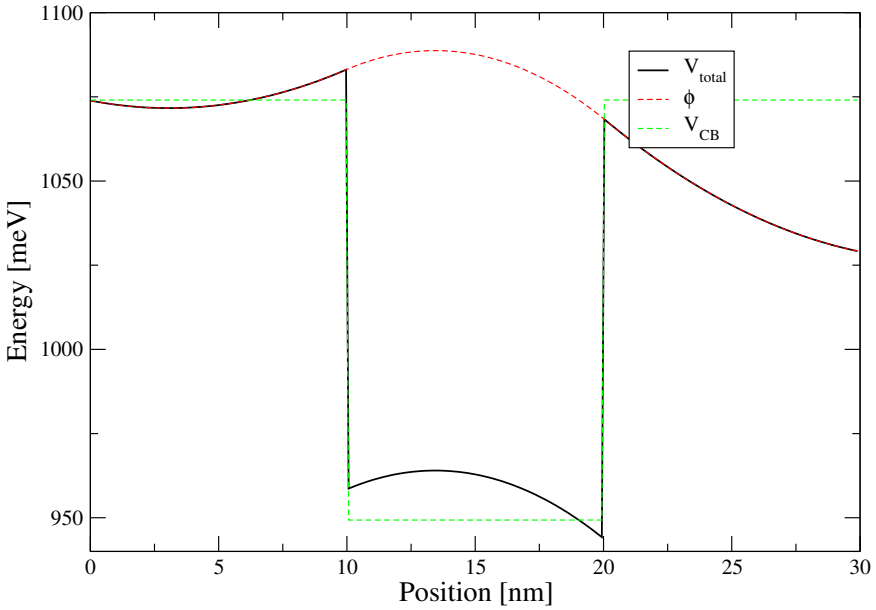

Figure 2.9: The Poisson potential, ϕ, for a 10 nm quantum well in the GaAs/$Al_{0.15}Ga_{0.85}As$ material system with barriers which are n-type doped to 1×10^{18} cm^{-3}, assuming all donors are ionised and free-carriers are evenly distributed throughout the entire structure, with an electric field across the structure of 15 kVcm^{-1}. The total band-edge profile, V_{total}, is also shown which is the sum of ϕ and the conduction band edge, V_{CB}. (The Poisson potential has been offset by the maximum value of the conduction band edge for clarity.)

Mixed boundary conditions

Since heterostructures are periodic structures, it would be advantageous to be able to model the charge effects within them using cyclic, or Neumann boundaries, to remove any effect the boundaries have on the results. However, implementing Neumann boundary conditions, such that the differential of potential is equal at each boundary, results in a singular matrix which cannot be solved. Physically, the reason for the matrix being singular can be viewed as there not being enough information to determine the potential; applying Neumann boundaries effectively isolates the system from *the outside world*, and since potentials are additive, there is no way of determining the absolute value of potential within the system. Equivalently, it can be seen as the potential within the system being dependent upon the potential in the next period of the structure, and therefore itself, meaning the potential cannot be found.

In order to avoid a singular matrix whilst still implementing a Neumann boundary, mixed boundary conditions can be used whereby both the differential and absolute value of potential are fixed at either one or both of the boundaries. In the case of Poisson's equation, the differential and absolute value cannot be fixed simultaneously at both boundaries as this results in the same singular matrix as with Neumann boundary conditions. (Physically this can be seen as forcing the potential drop across some charge whilst also fixing the electric field to be equal at each boundary — something which is not physically possible for charge distributions which are not antisymmetric around the centre of the system.) Instead, the differential may be fixed at both boundaries whilst the absolute value is only fixed at one boundary, giving the boundary conditions

$$\begin{aligned} \left.\frac{\partial\phi(z)}{\partial z}\right|_{i=1} &= \left.\frac{\partial\phi(z)}{\partial z}\right|_{i=N}, \\ \phi(z)|_{i=0} &= 0. \end{aligned} \tag{2.34}$$

Applying the finite difference approximation,

$$\frac{\phi_{i=1} - \phi_{i=0}}{\delta z} = \frac{\phi_{i=N+1} - \phi_{i=N}}{\delta z}, \tag{2.35}$$

and removing the redundant factor of δz gives the relation,

$$\phi_{i=1} - \phi_{i=0} = \phi_{i=N+1} - \phi_{i=N}. \tag{2.36}$$

Now substituting in ϕ_0,

$$\phi_{i=N+1} = \phi_{i=1} + \phi_{i=N}, \tag{2.37}$$

which by inserting into (2.26) gives a complete set of equations to solve:

$$\begin{aligned}
\frac{1}{\delta z}\left\{\varepsilon_{i+\frac{1}{2}}\phi_{i+1} - (\varepsilon_{i+\frac{1}{2}} + \varepsilon_{i-\frac{1}{2}})\phi_i\right\} &= -\rho_i, \qquad \text{for } i = 1; \\
\frac{1}{\delta z}\left\{\varepsilon_{i+\frac{1}{2}}\phi_{i+1} - (\varepsilon_{i+\frac{1}{2}} + \varepsilon_{i-\frac{1}{2}})\phi_i + \varepsilon_{i-\frac{1}{2}}\phi_{i-1}\right\} &= -\rho_i, \qquad \text{for } i = 2, 3, \ldots, N-1; \\
\frac{1}{\delta z}\left\{\varepsilon_{i+\frac{1}{2}}(\phi_{i=1} + \phi_{i=N}) - (\varepsilon_{i+\frac{1}{2}} + \varepsilon_{i-\frac{1}{2}})\phi_i + \varepsilon_{i-\frac{1}{2}}\phi_{i-1}\right\} &= -\rho_i, \qquad \text{for } i = N;
\end{aligned} \tag{2.38}$$

where the effective permittivity may be assumed constant across the boundaries of the simulation domain.

Rearranging into a matrix equation of the form (2.27), the matrix to solve, A, takes the form

$$\mathsf{A} = \begin{pmatrix}
b_1 & a_1 & 0 & \ldots & 0 \\
a_1 & b_2 & a_2 & \ldots & \vdots \\
0 & \ddots & \ddots & \ddots & 0 \\
\vdots & \ldots & a_{N-2} & b_{N-1} & a_{N-1} \\
c & \ldots & 0 & a_{N-1} & b_N
\end{pmatrix}, \tag{2.39}$$

where

$$\begin{aligned}
a_i &= \frac{1}{\delta z}\varepsilon_{1+\frac{1}{2}} \qquad \text{for } i = 1, 2, \ldots, N-1; \\
b_i &= \begin{cases} -\frac{1}{\delta z}(\varepsilon_{i+\frac{1}{2}} + \varepsilon_{i-\frac{1}{2}}) & \text{for } i = 1, 2, \ldots, N; \\ -\frac{1}{\delta z}\varepsilon_{i-\frac{1}{2}} & \text{for } i = N; \end{cases} \\
c &= \frac{1}{\delta z}\varepsilon_{i=\frac{1}{2}}.
\end{aligned} \tag{2.40}$$

Since the matrix to solve has a non-zero element in the bottom left corner, the simplest method of solution is to use a dense matrix digitalisation routine. This has the disadvantage of having very poor scalability in terms of execution time as the number of operations required increase roughly quadratically with the number of points in the simulation domain. The other option is utilise sparse matrix digitalisation routines, such as the PARDISO solver, which will be more efficient but are often more complicated to implement. Alternatively, an optimised algorithm may be developed for solving this specific form of matrix, as detailed in appendix B. This matrix is particularly well suited to this sort of optimisation process as it's form is so close to a tridiagonal matrix as shown in the proceeding section.

As solving Poisson's equation with mixed boundary conditions for a symmetric (about the centre of the simulation domain) charge distribution will produce a Poisson potential much the same as for Dirichlet boundaries, i.e. with no potential drop across the structure, the same 10nm quantum well as in 2.8 is simulated with only the right barrier being doped.

The resulting Poisson potential is shown in 2.10 where there is a potential drop across the structure of 142.2 meV.

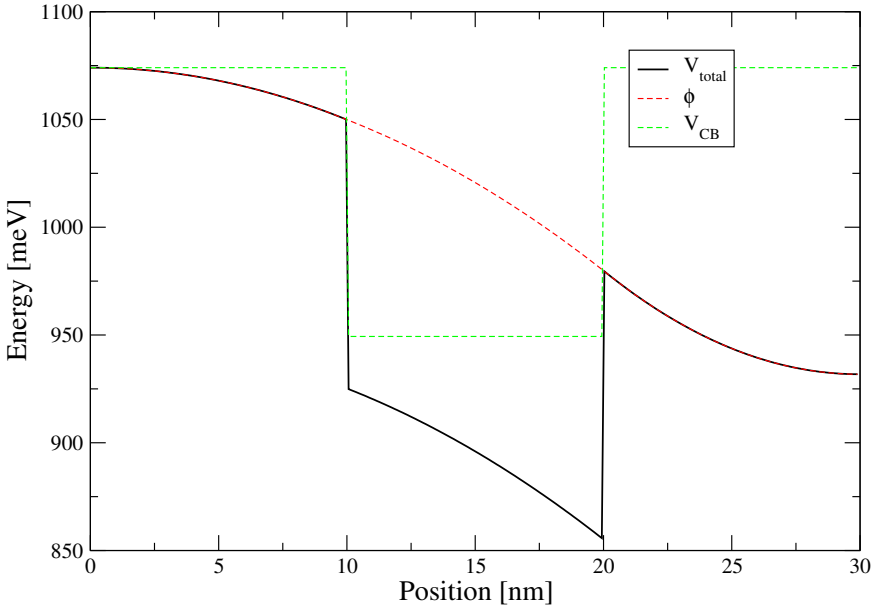

Figure 2.10: The Poisson potential, ϕ, for a 10 nm quantum well in the GaAs/$Al_{0.15}Ga_{0.85}As$ material system with the right barrier being n-type doped to $1 \times 10^{18}\,\mathrm{cm}^{-3}$, assuming all donors are ionised and free-carriers are evenly distributed throughout the entire structure. The total band-edge profile, V_{total}, is also shown which is the sum of ϕ and the conduction band edge, V_{CB}. (The Poisson potential has been offset by the maximum value of the conduction band edge for clarity.)

Optimised algorithm for solving mixed boundary condition

Inserting the form of matrix (2.39) in to the optimisation routine given in appendix B results in the following algorithm. As with the Thomas algorithm used for tridiagonal matrices, this requires both a forward and backward sweep to diagonalise the matrix. The forward sweep requires the values of the diagonal elements of A, b_i, and the values on the RHS of (2.27), ρ_i, to be given new values, b_i' and ρ_i' respectively, as well as a third array of values stored which

shall be labelled z'_i. These take the values

$$b'_i = \begin{cases} b_i & \text{for } i = 1; \\ b_i - \frac{a_i a_{i-1}}{b'_{i-1}} & \text{for } i = 2, 3, \ldots, N; \end{cases}$$
$$\rho'_i = \begin{cases} \rho_i & \text{for } i = 1; \\ \rho_i - \frac{a_i L'_{i-1}}{b'_{i-1}} & \text{for } i = 2, 3, \ldots, N-1; \end{cases} \tag{2.41}$$
$$z'_i = \begin{cases} 1 & \text{for } i = 1; \\ -z_{i-1}\frac{a_{i-1}}{b_{i-1}} & \text{for } i = 2, 3, \ldots, N-1; \end{cases}$$

This forward sweep reduces the matrix A to its upper triangle form, with the RHS of (2.27) now being ρ'. Finally, the problem may be solved via back substitution:

$$\phi_i = \begin{cases} -\frac{\rho'_i}{b'_i} & \text{for } i = N; \\ \frac{-\rho'_i - a_i \phi_{i+1}}{b'_i} & \text{for } i = N-1, N-2, \ldots, 1; \end{cases} \tag{2.42}$$

Figure 2.11 shows a comparison of the execution time of the optimised mixed algorithm and the dense matrix solver from LAPACK [47] for a range of different sized discretisation grids. Since the dense matrix solver requires roughly N^2 operations to solve the problem, where N is the number of grid points, it's execution time scales very poorly. The optimised algorithm on the other hand, which requires roughly $2N$ operations to solve, fares much better with a runtime which is of the order of 100 ms. The inset of figure 2.11 shows the percentage error of the optimised algorithm taking the dense matrix solution as the *correct* solution, which is of the order one would expect the numerical error to be when using double precision variables, therefore the solutions are the same down to numerical precision.

Object orientation of Poisson solver

Since a solution to Poisson's equation is needed so regularly within simulations semiconductor heterostructures, it is useful to have a well implemented, convenient method for solving Poisson's equation. Clearly, placing the code for solving Poisson's equation within some easily callable function is the simplest way of achieving this, however the concept of object orientation provides a method to fully encapsulate the solution. This avoids the unnecessary passing of variables between functions of the Poisson solver by effectively hiding the inner working of the solver from the user, reducing the amount of misuse/user error when using the Poisson solver.

The object orientated solution of Poisson's equation creates the matrix to diagonalise within the constructor of the Poisson solver class, named PoissonSolver. The only variables

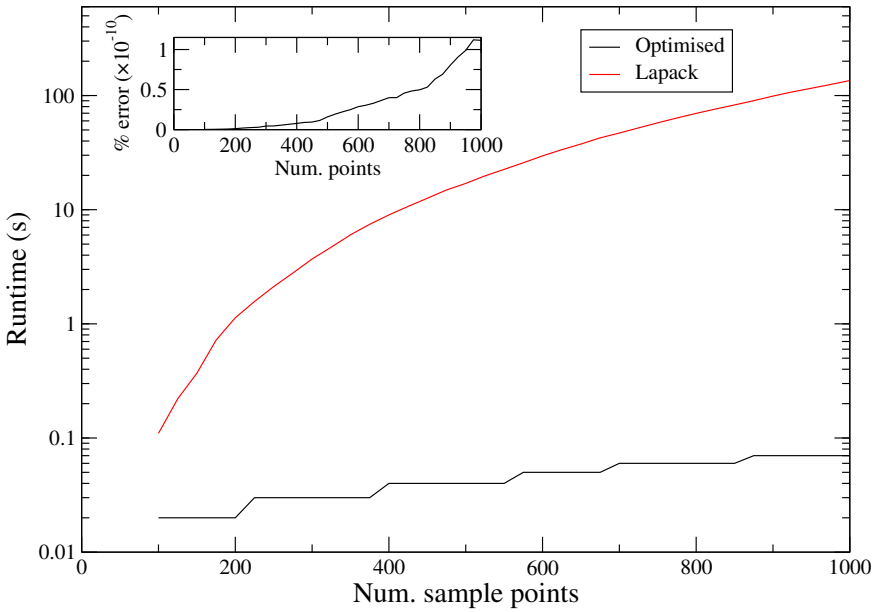

Figure 2.11: Comparison of runtimes of the Lapack dense matrix solver and the optimised algorithm for solving Poisson's equations with mixed boundary conditions. (Inset) Percentage error of optimised algorithm solution taking the Lapack solution to be correct.

which need to be passed to construct the matrix are the spatial discretisation step, δz, and the spatially varying effective permeability stored as an array of the values at each grid point, from which the number of grid points can be easily found. To solve Poisson's equation by diagonalising the matrix, a 'solve' member function may be called, to which only the spatially varying charge distribution needs to be passed, and the potential is returned from this function. Whats more, since linear algebra libraries such as LAPACK [47] have the functionality to diagonalise the matrix and perform the back substitution to find the solution in different functions, the matrix diagonalisation (where the majority of the time spent solving Poisson's equation is spent) may be done within the class constructor ather than the solve member function. Therefore if Poisson's equation is to be solved multiple times for the same structure but with different charge distribution (such as in the self-consistent Schrödinger-Poisson solution 2.3, the matrix only needs to be diagonalised once, making the solution more efficient. This has an even bigger effect for larger simulation domains since the matrix diagonalisation is slower, such as in the two-dimensional Poisson solution 3.2.3.

The user interface for the object orientated Poisson solver is

```
#include "PoissonSolver.h"

int main()
{
    // Define the simulation domain and create an array
    // of effective permeability, eps.
    double dx = 1e-10;
    vector<double> eps = getEps();

    // Create instance of PoissonSolver class.
    // Diagonalisation is done here so it doesn't need
    // to be done anywhere else.
    PoissonSolver poisson(dx, eps);

    // Solve for one charge distribution, rho_1
    vector<double> rho_1 = getRho();
    vector<double> phi_1 = poisson.solve(rho_1);

    // Solve for another charge distribution, rho_2
    vector<double> rho_2 = getAnotherRho();
    vector<double> phi_2 = poisson.solve(rho_2);

    // Do something with computed potentials...

    return EXIT_SUCCESS;
}
```

This clearly hides all implementation form the user stopping any unintentional editing of the implementation of the Poisson solver.

2.3 Self consistent Schrödinger-Poisson solutions

As discussed previously, localised charge within semiconductor heterostrucutres, coming from both bound, dopant ions and free carriers, cause a *band-bending* of the conduction band. However, since the spatially varying charge distribution of free carriers is dependent upon this

band-bending (since a change in the shape on the conduction band will shift the positional expectation values of the wavefunctions), a direct solution to find the exact form the conduction band edge including charge effects is not easily obtained. Instead, a self-consistent solution is opted for as the implementation of such an algorithm is simple and can be made to converge rapidly (i.e. within a few, < 5, iterations). Such a self-consistent solution may take the form shown in 2.12.

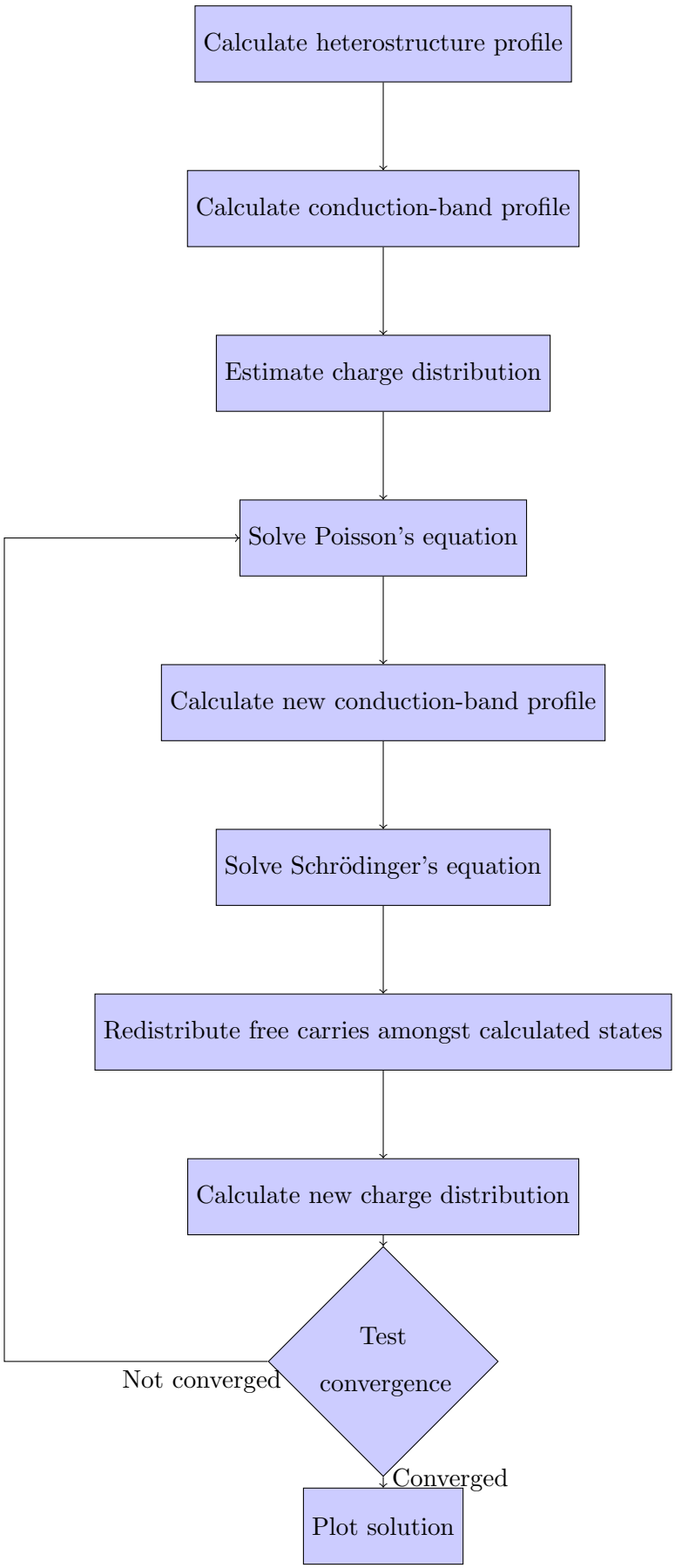

Figure 2.12: Flow diagram of self-consistent Schrödinger-Poisson solution.

While many of the blocks within 2.12 are self-explanatory, there are several which may differ depending on the implementation. These are discussed in detail here:

Calculate heterostrucutre profile:

The starting point for any simulation of a semiconductor heterostructure will be a definition of the structure, usually in the form a list of layer thickness, the alloy fraction within each of those layers and the positions of any dopants (if any). The purpose of this block is to convert that information into a spatially varying set of arrays containing useful information for the proceeding blocks; namely, the refractive index and doping profile. (Technically the conduction-band profile is also calculated here, however because it so important within the Schrödinger-Poisson loop it has been kept separate for clarity.)

Estimate charge distribution:

As discussed in 2.2, the simplest approximation to the distribution of free carriers is to assume that they are distributed evenly across the structure. This approximation is used within this block and gives a reasonable initial guess as to the total conduction-band profile without having to solve Schrödingers equation first.

Calculate new conduction-band profile:

The total conduction band profile is the sum of the unperturbed conduction-band profile and the potential from Poisson's equation, i.e.

$$V_{\text{total}} = V_{\text{CB}} + \phi.$$

However, for systems which have very large Poisson potentials (which may come from either very high doping concentrations or from in-built potentials such as those in GaN where strain induces a large potential within each layer), simply adding the potentials together can produce a self-consistent solutions which is not stable but either oscillates around the solution or diverges from the solution entirely. For such systems, it is beneficial to *damp* the rate of convergence by altering the method by which the Poisson potential is calculated such that, at each iteration, i,

$$V_{\text{Poisson}}|_i = (1-\alpha)\phi_i - \alpha V_{\text{Poisson}}|_{i-1},$$

and total potential now becomes becomes

$$V_{\text{total}} = V_{\text{CB}} + V_{\text{Poisson}}.$$

While this does increase the number of iterations required for the self-consistent loop to converge, it is need to ensure stability for certain systems. The effects of damping are shown below.

Redistribute free carriers amongst calculated states:

Once Schrödingers equation has been solved for the updated conduction-band profile, the free carriers within the system may be redistributed amongst the calculated states and there total spatial profile calculated. Deciding the population of each state, however, is a decidedly non-trivial task. The simplest solution is to distribute free charge amongst all states evenly, but even this isn't straightforward as it must be decided how many states should be populated as the most energetic states within a quantum well system are often not populated. A more accurate solution would be a temperature dependent distribution of carriers between states, however this requires the *quasi*-Fermi level to be calculated for each state as well as knowledge of the substrate and electron temperatures of the systems. (The electron temperature in particular is difficult to obtain, as described briefly in 5.2 and in more detail in [41], as it requires a rate equation model of the system which is part of self-consistent loop outside of the Schrödinger-Poisson loop.) The solutions within this section redistribute carriers evenly across all states within the quantum well system. Once the carriers have been distrusted amongst calculating the total spatial-varying charge distribution simply involves multiplying the population in each state by the states probability density and summing over all states for each point.

Test convergence:

Whether the solution has converged is a difficult problem to test. The traditional test is to check the value of one particular state energy (usually the ground state) from one iteration to the next; if the energies are within some predefined limit then solution is assumed to have converged. However, if the state being examined happens to be spatially confined to one region of the system, as is often the case with more complicated structure such as QCLs, this method provides no way of determining weather states within other regions of the structure have converged. Furthermore, if the solution does not converge then the loop will never exit and become infinite. Not only is this bad from an implementation point of view, but could also be problematic in larger calculations which may be more automated and so the infinite loop may not be spotted executing. The simplest solution to this problem is to set the loop to execute a set number of times and assume the solution will have converged by that point.

While this is a crude solution, it is often the most attractive as the convergence can normally be checked *by-eye* once it has completed and the number of iterations can be set to more than is required to ensure the solution has converged.

2.3.1 Results

Results are presented for a self-consistent solution of a 10 nm quantum well in the $Al_{0.3}Ga_{0.7}As/GaAs$ material system. Two doping are considered, well doping and barrier doping, both with a carrier concentration of $5\times10^{12}\,cm^{-2}$. This unusually high carrier concentration is used to produce a pronounced band-bending effect.

Figure 2.13 shows the results for the well doped quantum well after the self-consistent loop has converged, showing that considerable band-bending occurs with such high doping. To examine the performance of the self-consistent loop, as well as the effect of damping on the speed of convergence, the energy of ground state is plotted against the number of iterations. This is shown in figure 2.14. While the undoped solution does overshoot slightly within the first few iteration, it does oscillate around the solution and therefore damping (especially beyond $\alpha = 0.2$) only acts to slow convergence. The change in the conduction band edge with each iteration is shown in 2.15. This shows that the initial estimate for the charge density is quite far off the converged solution, as the converged solution for the total potential is closer to the unperturbed state than that of iteration 1. This is due to the barriers of the quantum well being thicker than the quantum well itself; since the majority of the probability density of all states will exist within the well, the assumption that free carriers will be evenly distributed is not correct in this case. This assumption does not stop the solution converging however, and for real systems which are much larger and typically have barriers which are very thin, and therefore have states with higher probability density inside the barriers, this assumption will be more valid.

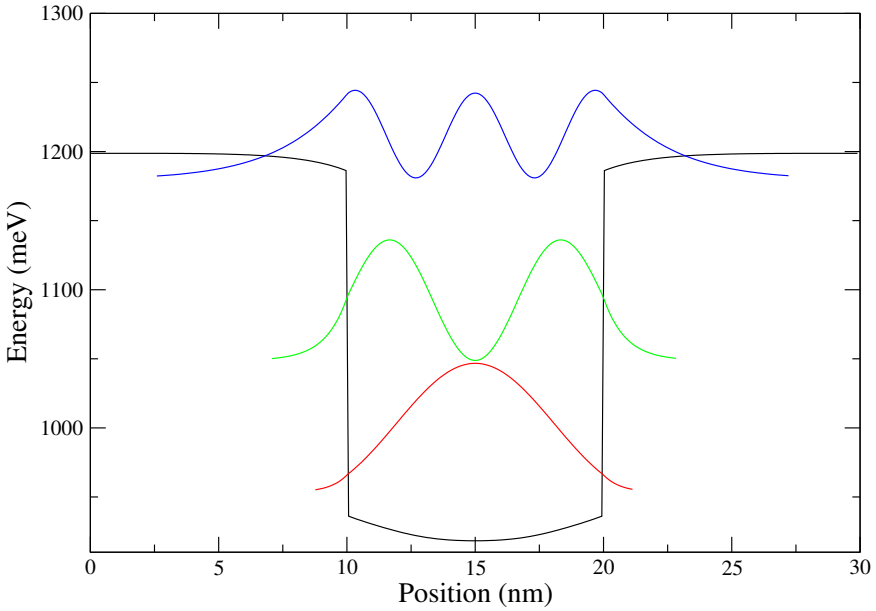

Figure 2.13: Self-consistent Schrödinger-Poisson solution for a 10 nm quantum well in the $Al_{0.3}Ga_{0.7}As/GaAs$ material system with dopant placed throughout the entire well such the doping density of the entire structure is $5\times10^{12}\,cm^{-2}$.

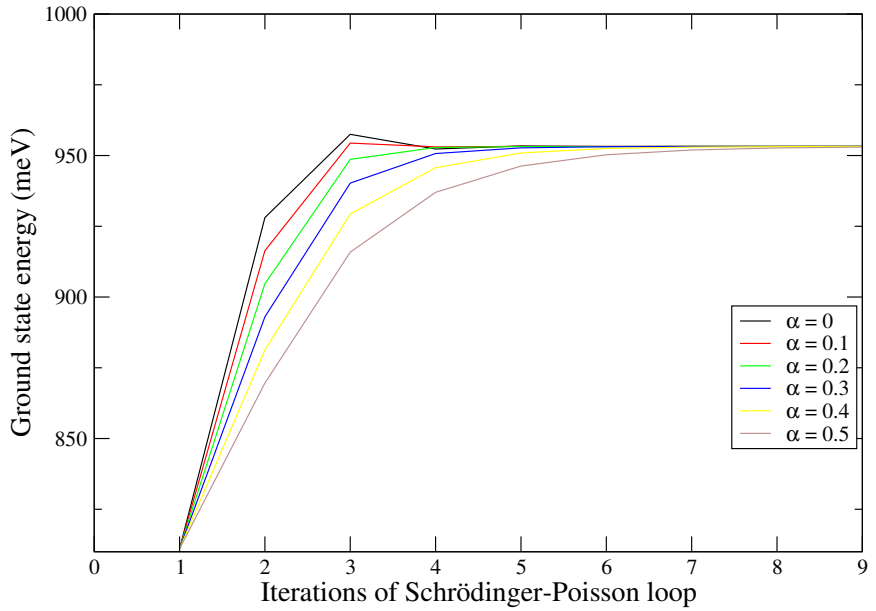

Figure 2.14: Evolution of the energy of the ground state of the single quantum well under consideration with iterations of the Schrödinger-Poisson loop for various values of the damping parameter α.

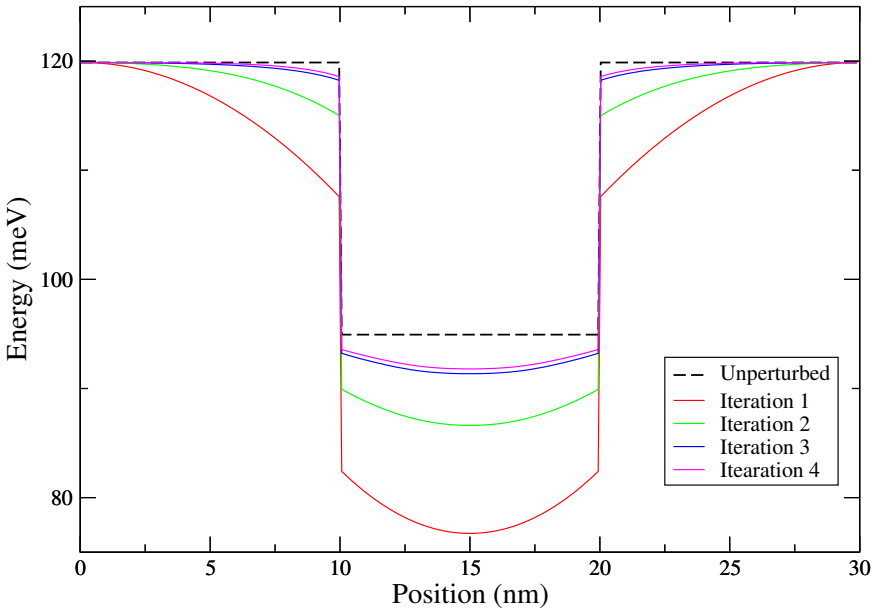

Figure 2.15: The change of the condition band edge for the single quantum well under consideration.

Figure 2.16 shows the resulting converged solution for the same quantum well but with doping placed within the barriers. The band-bending is even more pronounced here as the dopant ions and free carriers are more spatially separated that in the well-doped case (since, as discussed previously, the majority of free carriers will exist within the well). Figure 2.17 shows the evolution of the ground state energy with each iteration of the self-consistent loop. In this case, the undamped solution does oscillate before converging and therefore adding a small amount of damping does increase the speed of convergence. The change in the conduction band edge with each iteration is shown in 2.18.

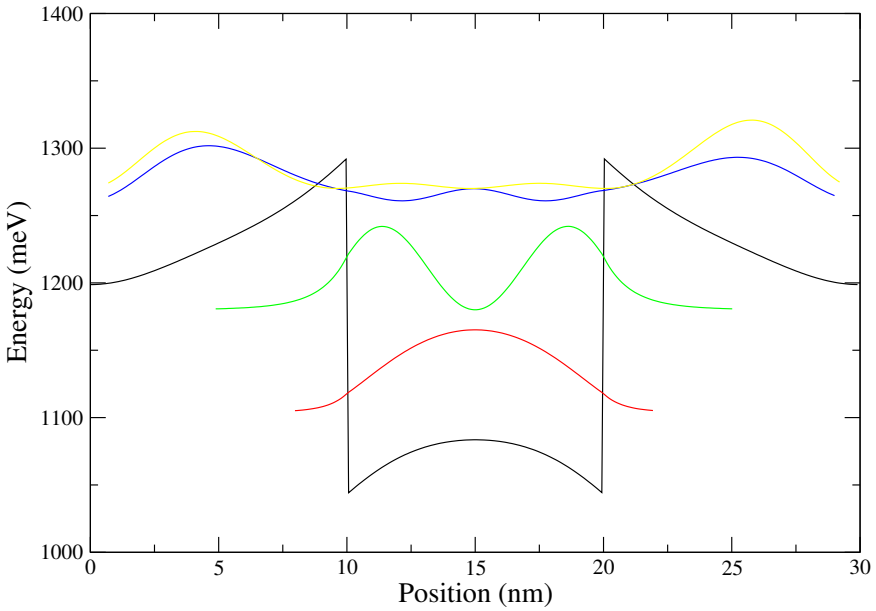

Figure 2.16: Self-consistent Schrödinger-Poisson solution for a 10 nm quantum well in the $Al_{0.3}Ga_{0.7}As/GaAs$ material system with dopant placed throughout each barrier such the doping density of the entire structure is $5\times10^{12}\,cm^{-2}$.

2.4 Conclusions

Within this chapter a model for finding the quantum states that exist within an arbitrary, one-dimensional quantum well system has been presented. This allows the relative energies and envelope functions of confined states within a semiconductor heterostructure to be determined, from which a device's electronic structure may be analysed such that some desirables properties may be designed in to a structure, e.g. a population inversion between the upper and lower lasing level for a QCL. Such a model is invaluable in the design of complex het-

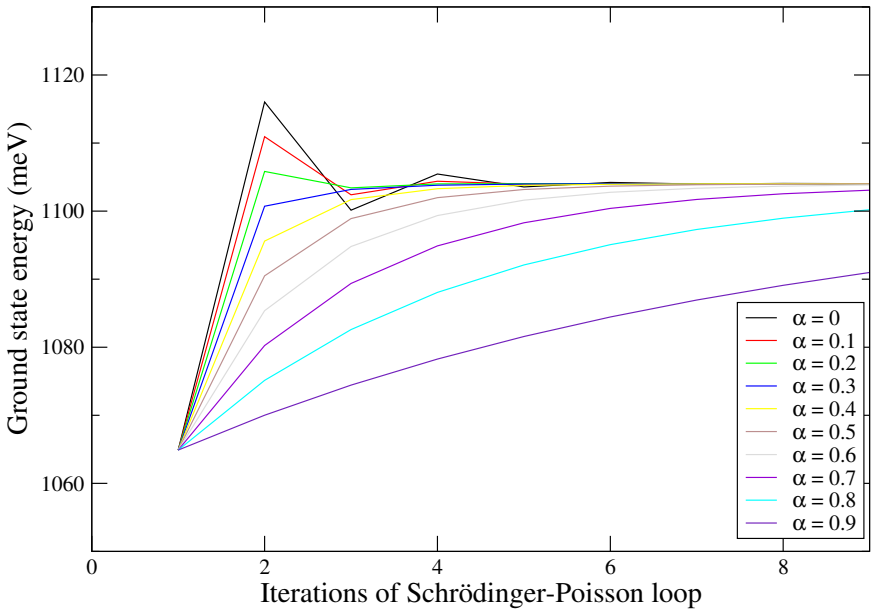

Figure 2.17: Evolution of the energy of the ground state of the single quantum well under consideration with iterations of the Schrödinger-Poisson loop for various values of the damping parameter α.

erostructure devices, such as QCLs, as it allows predictions about simple properties to be made directly from its results, for example a QCLs emission frequency, and can also be used as basis for a more in-depth analysis of the electronic structure, for example to examine the population of each state or to find the complex refractive index of a QCL gain medium as discussed in chapter 5.

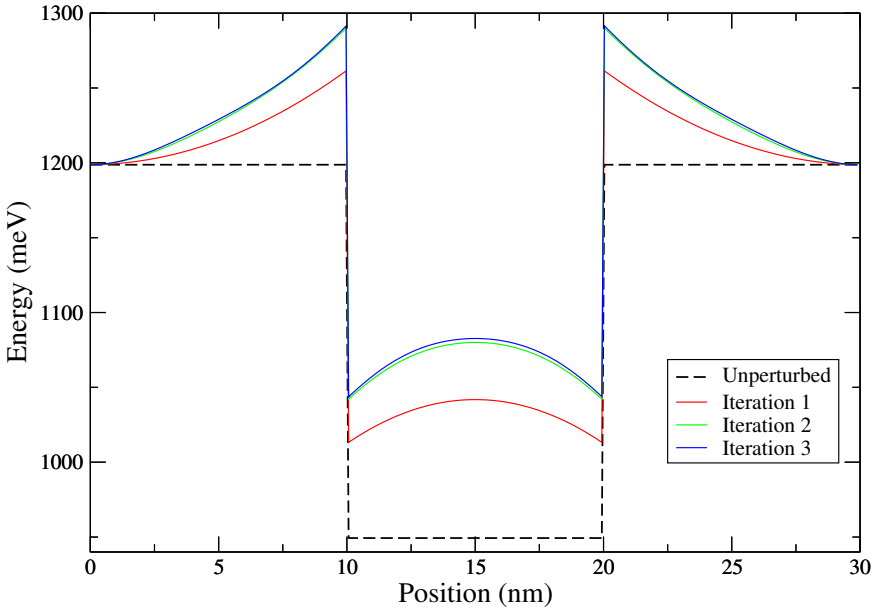

Figure 2.18: The change of the condition band edge for the single quantum well under consideration with $\alpha = 0.3$.

Chapter 3

Acoustic wave propagation in piezoelectric crystals

In order to determine how a surface acoustic wave (SAW) propagating through a quantum cascade laser (QCL) will effect the electronic properties of the active region, the form of the of that acoustic wave must be known. Since the structure of the SAW-QCL device is likely to be far from trivial as discussed in chapter 1 (i.e. with ridge structures and different layers of materials [36]), the method for determining the form of this acoustic wave must be a general model, flexible enough to accommodate any structural features required.

With these goals in mind, this chapter is structured as follows: First, the acoustic wave equations of motion are derived from first principles by considering the form deformation induced by an acoustic wave. Since this section contains numerous tensor equations, it makes heavy use of the Einstein summation convention, details of which can be found in appendix D. Secondly, the finite-difference time-domain scheme which is used to solve the equations of motion is presented, describing exactly how a solution is obtained. Finally, perfectly-matched-layer boundary conditions are derived for the model, which play a crucial role in stopping artificial reflections from the boundaries of the simulation domain interfering with the physics being investigated.

3.1 Stress and strain in crystalline materials

In order to gain a deeper understanding of acoustic wave propagation in crystalline materials, one must first understand how the deformations induced by an acoustic wave will alter a crystal on a bulk scale. Only when an accurate description of this deformation has been

grasped may the equation governing acoustic wave propagation be derived. Furthermore, deriving the equations of motion from first principles allows not only the accurate application of those equations and a better understanding the results coming from there solution, but also assists in the derivation of boundary conditions and other areas where the equations must be altered (such as in 3.3, 4.2.1 and 5.1.3).

With this is mind, this section begins by describing how stress and strain may be described from some arbitrary deformation. Stress and strain are then related to produce Hooke's law and the piezoelectric effect is added before the equations of motion are derived. Finally, tensor rotation is considered as this is required for applying the equations of motion to a set of axes in an arbitrary direction.

3.1.1 Tensor representation of stress and strain

When considering mechanical forces acting on a crystal lattice, it is necessary to consider the effect these forces have on the crystal unit cell. Under some resultant force or *stress*, the unit cell will deform and therefore become *strained*. If the unit cell returns to its original shape once the force has been removed then the crystal is defined as being elastic. These elastic deformations may be characterised by defining three orthogonal unit vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, such that these vectors lie along three axes of a Cartesian coordinate system, x_1, x_2 and x_3. Under some arbitrary deformation, the three unit vectors will be transformed into three new vectors $\mathbf{x}'$, $\mathbf{y}'$ and $\mathbf{z}'$, which may be defined as [50]

$$\begin{aligned} \mathbf{x}' &= (1+e_{11})\mathbf{x} + e_{12}\mathbf{y} + e_{13}\mathbf{z} \\ \mathbf{y}' &= e_{21}\mathbf{x} + (1+e_{22})\mathbf{y} + e_{23}\mathbf{z} \\ \mathbf{z}' &= e_{31}\mathbf{x} + e_{32}\mathbf{y} + (1+e_{33})\mathbf{z} \end{aligned} \tag{3.1}$$

where the coefficients e_{ij} are dimensionless and completely define the transformation of the three vectors, and therefore the deformation. For elastic deformations these coefficients are always $\ll 1$.

By considering the displacement induced by an arbitrary deformation, the coeffcients e_{ij} may be defined in terms of displacement. If the position vector of an atom is defined as $\mathbf{r} = \mathbf{x}+\mathbf{y}+\mathbf{z}$ such that it sits in the top far corner of a unit cell before deformation, it follows that its position after deformation will be $\mathbf{r}' = \mathbf{x}' + \mathbf{y}' + \mathbf{z}'$. Therefore the total displacement of the atom, $\mathbf{u}$, will be [50]

$$\mathbf{u}(\mathbf{r}) = \mathbf{r}' - \mathbf{r} = u_1(\mathbf{r})\mathbf{x} + u_2(\mathbf{r})\mathbf{y} + u_3(\mathbf{r})\mathbf{z} \tag{3.2}$$

where u_1, u_2 and u_3 are the individual components of the displacement aligned along the x_1, x_2 and x_3 axes respectively.

This relationship between the two vectors $\mathbf{u}$ and $\mathbf{r}$ may be written more succinctly by using a second-rank tensor [51],

$$\mathbf{u} = [\mathsf{e}]\mathbf{r} \tag{3.3}$$

where the second-rank tensor $[\mathsf{e}]$ is made up of the coefficients e_{ij} from (3.1), and has the form

$$e_{ij} = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix}. \tag{3.4}$$

This allows $[\mathsf{e}]$, and therefore the deformation, to be defined by the displacement induced on some arbitrary point,

$$e_{ij} = \frac{\Delta u_i}{\Delta r_j} \qquad \text{for } i, j = 1, 2, 3, \tag{3.5}$$

where Δu_i is the contribution to magnitude of the component of $\mathbf{u}$ along the x_i axis from the deformation e_{ij} and Δr_j is the magnitude of the component of $\mathbf{r}$ along the x_j axis. Since $\mathbf{r}$ has an arbitrary magnitude, making it arbitrarily small, i.e. $\Delta r_j \to \partial r_j$, will change the fraction in (3.5) to a differential. This is equivalent to making the assumption that the material being deformed is homogeneous, and turns out to be incredibly useful as it allows the displacement of any point within such a homogeneous material to be found for any arbitrary deformation. Making this assumption gives

$$e_{ij} = \lim_{\Delta r_j \to 0} \frac{\Delta u_i}{\Delta r_j} = \frac{\partial u_i}{\partial r_j} \qquad \text{for } i, j = 1, 2, 3, \tag{3.6}$$

At this point it is important to realise that while $[\mathsf{e}]$ is a complete description of the displacement of any point within a material, it does not describe the strain. This can be seen be considering what happens to $[\mathsf{e}]$ when a translation (or rotation) is applied. Since there will be displacement to every point (not at the center of rotation), $[\mathsf{e}]$ will be non-zero everywhere (except this point) despite the fact that the material has not been deformed and therefore there is no strain. However, since any arbitrary translation or rotation will produce a purely anti-symmetric tensor, and the fact that any second rank tensor may be defined as the sum of a symmetric and anti-symmetric tensor, i.e.

$$e_{ij} = \epsilon_{ij} + \varpi_{ij} \qquad \text{for } i, j = 1, 2, 3, \tag{3.7}$$

where $[\epsilon]$ and $[\varpi]$ are the symmetric and anti-symmetric tensors respectively, the strain, $[\epsilon]$,

may be defined as the purely symmetric part of [e],

$$\epsilon_{ij} = \frac{1}{2}(e_{ij} + e_{ji}) \qquad \text{for } i, j = 1, 2, 3. \tag{3.8}$$

A more rigorous proof of this derivation of strain is beyond the scope of this thesis but may be found in reference [51].

For completeness, the full strain tensor is then

$$[\epsilon] = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{31} \\ \epsilon_{12} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{23} & \epsilon_{33} \end{bmatrix} = \begin{bmatrix} e_{11} & \frac{1}{2}(e_{12} + e_{21}) & \frac{1}{2}(e_{13} + e_{31}) \\ \frac{1}{2}(e_{12} + e_{21}) & e_{22} & \frac{1}{2}(e_{23} + e_{32}) \\ \frac{1}{2}(e_{13} + e_{31}) & \frac{1}{2}(e_{23} + e_{32}) & e_{33} \end{bmatrix}. \tag{3.9}$$

Since each unique component of strain must be induced by a unique component stress, it may be deduced that stress must also be represented by a symmetric rank-two tensor. In order to define this stress tensor, the various ways in which stress can be applied to a unit cell are considered. As stress is a force per unit area, stress applied to the unit cell can be thought of as an instantaneous force applied equally to one entire face of the unit cell. Clearly, the three diagonal elements of the stress tensor, σ_{ii} for $i = 1, 2, 3$, will refer to a force applied perpendicular to the face it is applied to, inducing either a compressive or tensile strain (depending on whether the force is pointing inwards or outward from the unit cell). This stress, and the strain which it induces, is labelled as *lateral*. The only other unique way of applying stress to the unit cell is to apply a force parallel to the face it is applied to, which corresponds to the other three unique elements of the stress tensor, σ_{ij} for $i \neq j$ with $i, j = 1, 2, 3$. This stress, and induced strain, are labelled as *shear*. Applying stress at any other orientation will simply produce some combination of lateral and shear stress. Figure 3.1 shows specifically which subscripts apply to which faces and axes with hidden faces having the same notation as their opposite face (since the stress tensor is symmetric). The full form of the stress tensor is then

$$[\sigma_{ij}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{31} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{23} & \sigma_{33} \end{bmatrix}. \tag{3.10}$$

3.1.2 Hooke's Law

Hooke's law states that stress is directly propositional to strain, i.e.

$$\sigma = \mathsf{C}\epsilon, \tag{3.11}$$

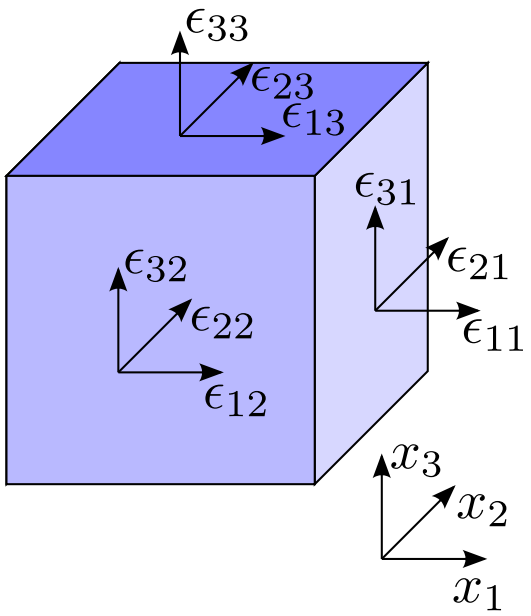

Figure 3.1: Orientation of the subscripts for both stress and strain (the orientation of subscripts is identical for both) for the visible faces of a cube. The non-visible faces will have identical subscripts to their opposing sides.

where C is the elastic stiffness constant which is a property specific to the material or system being stressed/strained. In order for (3.11) to be valid, the induced strain must be within the elastic limit, as defined in section 3.1.1. With the addition of Hooke's law, the elastic limit may be defined more generally. That is, (3.11) is only valid when the induced strain is small enough for the elastic stiffness constant to remain constant, i.e. remain within the linear regime. While higher order terms can be added to (3.11), or precisely in the definition of stress and strain, to account for deformations outside of this linear regime, such approximations are not necessary in this work. Instead, this limit of validity is considered wherever Hooke's law is applied.

As shown in section 3.1.1, stress and strain within crystalline materials may be represented by second-rank tensors. In order to use these in (3.11), the elastic stiffness constant must be a fourth-rank tensor (since rank-two tensors are related by rank-four tensors), i.e. $[\mathsf{C}]$. The tensor representation of Hooke's law within a crystalline material is then [51]

$$[\sigma] = [\mathsf{C}][\epsilon], \tag{3.12}$$

or using the Einstein summation convention (see appendix D)

$$\sigma_{ij} = C_{ijkl}\epsilon_{kl} \qquad \text{for } i, j, k, l = 1, 2, 3. \tag{3.13}$$

Matrix representation of tensors

As discussed in section 3.1.1, both the stress and strain tensors are symmetric having only six unique elements out of nine each. Therefore the elastic stiffness tensor will also be symmetric, having only thirty six unique elements out of a total eighty one. This level of symmetry within each tensor lends itself very well to representation by matrices, and is incredibly beneficial as it greatly reduces the complexity of the analysis when applying Hooke's law.

Representing tensors using matrices simply involves changing the notation used to represent each element. Since the stress and strain tensors have six unique elements, the first step is to label these with a single suffix, from 1 to 6, replacing the double suffix notation, as in table 3.1. Using this notation implies that the stress and strain tensors may be represented

Tensor notation	Matrix notation
11	1
22	2
33	3
23 or 32	4
31 or 13	5
12 or 21	6

Table 3.1: Conversion from tensor to matrix notation.

by vectors each with 6 elements, which must be related by a 6×6 matrix in order for (3.11) to hold. Such a matrix can be constructed from the elastic stiffness tensor by converting the first and second pair of suffixes into matrix notation, as in table 3.1, i.e.

$$C_{ijkl} \rightarrow C_{mn} \qquad \text{for } i,j,k,l = 1,2,3;\ m,n = 1,\ldots,6.$$

Care must be taken to ensure that this matrix representation is identical to the tensor representation. If the tensor product in (3.12) is considered, it can be seen that implied conversion to matrices described above does not produce an identical equation, since the unique elements are included the number of times they appear in the tensor representation but only once in the matrix representation. This can be rectified by introducing a factor of two when converting the shear strain components, therefore making the conversion of strain from tensor to matrix representation [51]

$$\epsilon_i = \epsilon_{ii} \qquad \text{for } i = 1,2,3,$$

$$\epsilon_i = 2\epsilon_{jk} \qquad \text{for } i = 4,5,6;\ j,k = 1,2,3, \text{ with } j \neq k,$$

or, representing the elements of $[\epsilon]$ in matrix notation

$$\begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{bmatrix} \rightarrow \begin{bmatrix} \epsilon_1 & \frac{1}{2}\epsilon_6 & \frac{1}{2}\epsilon_5 \\ \frac{1}{2}\epsilon_6 & \epsilon_2 & \frac{1}{2}\epsilon_4 \\ \frac{1}{2}\epsilon_5 & \frac{1}{2}\epsilon_4 & \epsilon_3 \end{bmatrix}.$$

This matrix representation of shear strain is sometimes referred to as *engineering shear strain*. The elements of stress do not require any scaling factor and, for completeness, the conversion from tensor matrix is

$$\sigma_i = \sigma_{jk} \qquad \text{for } i, j, k = 1, 2, 3,$$

or, representing the elements of $[\sigma]$ in matrix notation

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \rightarrow \begin{bmatrix} \sigma_1 & \sigma_6 & \sigma_5 \\ \sigma_6 & \sigma_2 & \sigma_4 \\ \sigma_5 & \sigma_4 & \sigma_3 \end{bmatrix}.$$

By expanding (3.13) for one element of $[\sigma]$, say σ_{11}, in both tensor and matrix notations it can be seen that they are now equivalent, i.e.

$$\begin{aligned} \sigma_{11} = {} & C_{1111}\epsilon_{11} + C_{1112}\epsilon_{12} + C_{1113}\epsilon_{13} \\ & + C_{1121}\epsilon_{21} + C_{1122}\epsilon_{22} + C_{1123}\epsilon_{23} \\ & + C_{1131}\epsilon_{31} + C_{1132}\epsilon_{32} + C_{1133}\epsilon_{33}, \end{aligned}$$

which becomes,

$$\begin{aligned} \sigma_1 = {} & C_{11}\epsilon_1 + \frac{1}{2}C_{16}\epsilon_6 + \frac{1}{2}C_{15}\epsilon_5 \\ & + \frac{1}{2}C_{16}\epsilon_6 + C_{12}\epsilon_2 + \frac{1}{2}C_{14}\epsilon_4 \\ & + \frac{1}{2}C_{15}\epsilon_5 + \frac{1}{2}C_{14}\epsilon_4 + C_{13}\sigma_3, \end{aligned}$$

in matrix notation.

The matrix representation of Hooke's law in crystalline materials may now be written out in full,

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46} \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56} \\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66} \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{pmatrix}, \tag{3.14}$$

or using the Einstein summation convention,

$$\sigma_i = C_{ij}\epsilon_j \qquad \text{for } i,j = 1,\ldots,6, \tag{3.15}$$

where ϵ may be expressed in terms of the displacement inside the material, as described in section 3.1.1,

$$\begin{aligned}
\epsilon_i &= \frac{\partial u_i}{\partial x_i} \qquad \text{for } i = 1,2,3,\\
\epsilon_4 &= \frac{\partial u_3}{\partial x_2} + \frac{\partial u_2}{\partial x_3},\\
\epsilon_5 &= \frac{\partial u_3}{\partial x_1} + \frac{\partial u_1}{\partial x_3},\\
\epsilon_6 &= \frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2}.
\end{aligned} \tag{3.16}$$

3.1.3 Including the piezoelectric effect

When a strain is induced within certain crystals, asymmetry of atomic positions within the unit cell can cause an imbalance in charge which generates an electric field. This is known as the *direct piezoelectric effect* [51]. Similarly, applying an electric field to a crystal which exhibits the piezoelectric effect will induce a electric displacement vector within the crystal causing it to become strained. This is known as the *converse piezoelectric effect* [51]. As stated in section 3.1.2, the stresses and strains considered in this work are assumed to be small and therefore within the elastic/linear limit. It can therefore be assumed that the piezoelectric effect will also be within this elastic limit, i.e. the strain is small enough the constant defining piezoelectricity will not changed under any arbitrary deformation. Since it is the displacement of atoms within the unit cell that induces a electric displacement vector, a relationship between the strain and the induced electric displacement vector is sought. This relationship will be linear, with the constant defining the piezoelectricity being a third-rank tensor, since it relates a second-rank tensor, strain, to a vector, the electric displacement vector, and will take the form [51]

$$\mathbf{D} = [\mathsf{e}][\epsilon] \tag{3.17}$$

where $\mathbf{D}$ is the electric displacement vector and $[\mathsf{e}]$ is the piezoelectric constant, or using the Einstein summation convention

$$D_i = e_{ijk}\epsilon_{jk} \qquad \text{for } i,j,k = 1,2,3, \tag{3.18}$$

where D_i is the electric displacement vector along the x_i axis. The strain may be related to the induced potential by remembering the divergence of the electric displacement vector

gives the charge distribution,

$$\nabla \cdot \mathbf{D} = \nabla \cdot [\mathsf{e}][\epsilon] = \rho. \tag{3.19}$$

This may be substituted into Poisson's equation to find the potential,

$$\nabla \varepsilon \nabla \phi = -\nabla \cdot [\mathsf{e}][\epsilon] \tag{3.20}$$

where ε is the permittivity and ϕ is the potential.

A relationship is now sought to relate an applied electric field to an induced stress. This can also be assumed to be linear and will also be a third-rank tensor since it is relating a vector and rank-two tensor. It can be shown thermodynamically, see [51], that this constant is in fact the transpose of the piezoelectric constant, which in the Einstein summation convention is equivalent to reversing the order of the suffixes. This simply allows an extra terms to be added to Hooke's law to include the piezoelectric effect, which will be negative since the direct effect must oppose the converse effect,

$$[\sigma] = [\mathsf{C}][\epsilon] - [\mathsf{e}]\mathbf{E}, \tag{3.21}$$

where is $\mathbf{E}$ is the electric field, or using the Einstein summation convention

$$\sigma_{ij} = C_{ijkl}\epsilon_{kl} - e_{mji}E_m \qquad \text{for } i,j,k,l,m = 1,2,3. \tag{3.22}$$

Since the solution of Poisson's equation gives a potential, it is often easier from an implementation point of view to include the potential in Hooke's law, which may be done very simply by remembering that the gradient of potential gives the negative electric field, and therefore Hooke's law becomes

$$[\sigma] = [\mathsf{C}][\epsilon] + [\mathsf{e}]\nabla\phi, \tag{3.23}$$

or using the Einstein summation convention

$$\sigma_{ij} = C_{ijkl}\epsilon_{kl} + e_{mij}\phi \qquad \text{for } i,j,k,l,m = 1,2,3. \tag{3.24}$$

For completeness, the piezoelectric form of Hooke's law may be written in matrix notation by converting the piezoelectric tensor into matrix notation. This is done by substituting the second two tensor suffixes to a single matrix suffix as in table 3.1 such that,

$$e_{ijk} = e_{il} \qquad \text{for } i,j,k = 1,2,3; \quad l = 1,\ldots,6, \tag{3.25}$$

which results in a 6×3 matrix of the form

$$\mathbf{e} = \begin{pmatrix} e_{11} & e_{12} & e_{13} & e_{14} & e_{15} & e_{16} \\ e_{21} & e_{22} & e_{23} & e_{24} & e_{25} & e_{26} \\ e_{31} & e_{32} & e_{33} & e_{34} & e_{35} & e_{36} \end{pmatrix}. \tag{3.26}$$

In Hooke's law the transverse of this matrix must be used and in the summation convention this becomes

$$\sigma_i = C_{ij}\epsilon_j + e^{\mathrm{T}}_{ik}\frac{\partial \phi}{\partial x_k} \tag{3.27}$$

3.1.4 Acoustic wave equations of motion

The defining property of an acoustic wave is a time varying displacement at each point within the medium the wave is travelling through. As already shown in section 3.1.2, displacement, or strain, within crystalline materials is induced by some stress. A relation is then sought between this time varying displacement and the instantaneous stress in the form of a wave equation to describe acoustic wave propagation. A similar relation is found in Newton's second law,

$$F = m\frac{\partial^2 S}{\partial t^2}, \tag{3.28}$$

where an instantaneous force, F, is related to the second-order time differential of displacement (from some origin), S, by the mass, m. There are two main differences between (3.28) and the wave equation which is sought. Firstly, the displacement of a point within a crystalline material is not measured from the origin but from where that point lies in the unstrained material, as in seciton 3.1.1. Therefore this displacement can be though of as a displacement per unit of volume. Secondly, there is no instantaneous force, but an instantaneous stress, i.e. a force per unit area. However, if the spatial differential of strain is taken to give a force per unit volume, then using the density (mass per unit volume) an analogous equation may be written down effectively describing Newton's second law per unit volume,

$$\nabla \cdot [\sigma] = \rho\frac{\partial^2 \mathbf{u}}{\partial t^2} \tag{3.29}$$

where ρ is the density and $\mathbf{u}$ is displacement, or using the matrix notation and Einstein summation convention,

$$\rho\frac{\partial^2 u_i}{\partial t^2} = \frac{\partial \sigma_{ij}}{\partial x_j} \qquad \text{for } i,j = 1,2,3. \tag{3.30}$$

This is the acoustic wave equation motion relating the time-dependent displacement to the instantaneous stress within a three-dimensional crystalline material. (Note that in (3.30) σ is in tensor rotation, as this is the only way that this equation can be written using the Einstein summation convention.) It is worth noting that (3.30) is incredibly general. Indeed, any form of acoustic wave propagation in a bulk crystalline material is described by this equation. In fact propagation in non-bulk materials can also be described by applying appropriate boundary conditions, such as with surface acoustic waves as in chapter 4.

3.1.5 Tensor rotation

Although not explicitly stated until now, the spatial axes of the equations of motion derived in 3.1.4 will be aligned along the crystal axes of the crystalline material being simulated unless the axes of tensor constants are rotated. This is because the axes of the tensor constants within the equations of motion are aligned to the crystal axes, or rather the constants are always stated with their axes aligned to the crystal axes, and the axes of the equations of motion are implicitly aligned along the tensor's axes. Since it is often necessary to align the axes of the equation of motion to some arbitrary direction, especially when the simulation domain is restricted to a quasi-three dimensional system as described in 3.2.1 where the direction of the restricted axis determines the direction within the crystal of the propagating wave, a method is required to rotate the tensor axes.

Since the method of rotation requires a transformation matrix to be defined for an arbitrary rotation, this is done first. Then the method of rotation is described which utilises the transformation matrices.

Transformation matrices for arbitrary rotations

Any arbitrary rotation which transforms some coordinate system defined by three orthogonal axes, Ox_1, Ox_2 and Ox_3, onto three new orthogonal axes, Ox_1, Ox_2 and Ox_3, may be defined by three distinct angles [51, p. 33]. This may be shown by considering a rotation which moves one of the three axes, Ox_1 say, to some new direction Ox_1'. The movement of Ox_1 to Ox_1' may be described fully by a rotation of Ox_1 around a plane on which it already lies followed by an elevation away from that plane. This rotation leaves the position of the other two axes undefined however, since the Ox_1 axis may be rotated further about itself. Therefore a third angle is required to describe the transformation of axes fully. These three angles can be thought of as rotations about each of the three orthogonal axes[1], therefore allowing any arbitrary rotation to be described by a combination rotations about each axis as shown in 3.2. Three transformation matrices are therefore sought to describe a rotation about each of the three axes so that a combination of these may be used to describe any arbitrary rotation.

At this point, it is worth noting about the convention for positive and negative angles in 3-dimensional space. In the present work the convention from [51] is used, that is on *right-handed* axes a positive rotation about one axis is defined as the direction of a right-handed screw motion along that axis. Since only rotations about one axis are considered here, this

[1] Although in fact the three angles can be taken about any three orthogonal vectors

definition need not be extended to include all planes of rotation.

In order to find the form of each transformation matrix, the effect of a rotation about one axis on the components of a rank 1 tensor, i.e. a vector, is examined since this is the simplest non-trivial case of tensor rotation.[2] Clearly the diagonal elements of the transformation matrix will be the cosine of the rotation angle, however the diagonal element which corresponds to the axis being rotated about will be one since the magnitude of the corresponding vector component will not change. The off diagonal elements will have the magnitude of the sine of the rotation angle, although to determine the sign of these elements one must consider the direction in which occurs. Furthermore, the elements of the column and row which have the diagonal element of one will be zero since the corresponding vector components will remain orthogonal to the axis being rotated about both before and after the rotation (because this axis does not change direction), and therefore will not affect nor be affected by the vector component in this direction.

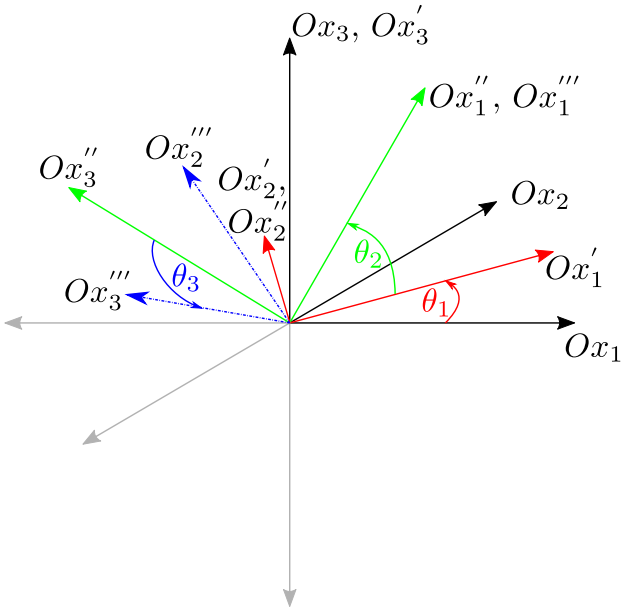

Figure 3.2: Arbitrary rotation of the three orthogonal axes, Ox_i, to some new position Ox_i''', via a rotation first about Ox_3 by an angle of θ_1 to Ox_i', followed by a rotation about Ox_2' by θ_2 to Ox_i'', and finally a rotation about Ox_1' by θ_3 to Ox_i'''.

Using the above conditions, the transformation matrices may be written down by considering the relation between some arbitrary vector, $\mathbf{v}$, and some resulting vector $\mathbf{v}'$, after a

[2]Rotations of rank 0 tensors, i.e. scalars, are simpler, but are trivial since a rotation of axes does not change the scalar value.

rotation about each of three axes. I.e., for a rotation about the axis Ox_1 by and angle θ

$$\mathbf{v}' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \mathbf{v}, \tag{3.31}$$

for a rotation about the axis Ox_2 by and angle θ

$$\mathbf{v}' = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \mathbf{v}, \tag{3.32}$$

and for a rotation about the axis Ox_3 by an angle θ

$$\mathbf{v}' = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{v}. \tag{3.33}$$

As the above transformation matrices may be combined to describe any arbitrary transformation, it is readily shown that the same transformation may be achieved by multiplying the transformation matrices together such that

$$\mathbf{v}' = \mathbf{a}_z \mathbf{a}_y \mathbf{a}_x \mathbf{v} \tag{3.34}$$

where $\mathbf{a}_x$, $\mathbf{a}_y$ and $\mathbf{a}_z$ are the transformation matrices for rotations about Ox_1, Ox_2 and Ox_3 respectively, in order to find a single, general matrix for any arbitrary rotation. However, because matrix multiplication is noncommutative, the order in which the rotations about Ox_1, Ox_2 and Ox_3 are performed will intrinsically set with any general transformation matrix. While this is not a problem so long as care is taken to insure that the total rotation is described in terms of rotations about axes in a specific order, some rotations are easier to describe by rotating about axes in different orders. Therefore, in order to avoid confusion about the precise order of rotation and the rigidity of having to perform the rotations in a specific order, multiplying the transformation matrices together to form a general matrix is avoided. Instead the three transformation matrices, $\mathbf{a}_x$, $\mathbf{a}_y$ and $\mathbf{a}_z$, for rotations about Ox_1, Ox_2 and Ox_3 respectively are given as

$$\mathbf{a}_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, \tag{3.35}$$

$$\mathbf{a}_y = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}, \tag{3.36}$$

$$\mathbf{a}_z = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.37}$$

Transformation of tensor constants

While the present work is concerned only with the rotation of tensors, the method of utilising a transformation matrix once it has been found for some arbitrary transformation (i.e. a rotation) is the same, and therefore will first be described in a general sense. In fact, the way in which a physical property transforms in terms of some transformation matrix *is* the defining property of a tensor, and a property is not considered a tensor if it does not transform as described below [51].

During a transformation, tensors are multiplied by a number of elements from the transformation matrix corresponding to their rank number, with the specific elements that are multiplied together changing dependent also upon the rank of the tensor. For rank 3 tensors, e.g. the piezoelectric constant, the transformation, using the Einstein summation convention, is

$$T'_{ijk} = a_{il}a_{jm}a_{kn}T_{lmn} \tag{3.38}$$

where T' is the tensor after transformation, $\mathbf{a}$ is the transformation matrix and T is the original tensor. For rank 4 tensors, e.g. the elastic constant, the transformation, using the Einstein summation convention, is

$$T'_{ijkl} = a_{im}a_{jn}a_{ko}a_{lp}T_{mnop} \tag{3.39}$$

where T' is the tensor after transformation, $\mathbf{a}$ is the transformation matrix and T is the original tensor.

As discussed during the derivation of the transformation matrices, the matrix for a rotation about each axis was kept separate to avoid confusion about which order the rotations take place. If an arbitrary rotation were to be performed as above, these three matrices would need to be multiplied together before the transformation could take place. However from an implementation point of view, it is easier to simply perform the rotation about each axis separately as this removes the need to multiply the transformation matrices together.

This will produce identical results to performing the transformation in one so long as care is taken that each rotation is performed separately, for example, firstly the transformation

$$T^x_{ijk} = a^x_{il} a^x_{jm} a^x_{kn} T_{lmn},$$

where T^x is tensor after a rotation about x and $\mathbf{a}^x$ is the transformation matrix for a rotation about x, followed by

$$T^y_{ijk} = a^y_{il} a^y_{jm} a^y_{kn} T^x_{lmn},$$

where T^y is tensor after a rotation about y and $\mathbf{a}^y$ is the transformation matrix for a rotation about y, and finally

$$T^z_{ijk} = a^z_{il} a^z_{jm} a^z_{kn} T^y_{lmn},$$

where T^z is tensor after a rotation about z and $\mathbf{a}^z$ is the transformation matrix for a rotation about z. Furthermore, many rotations will only require a rotation about one axis, such as when acoustic propagation in the [1,1,0] direction which requires only a rotation about x by 45°, and therefore no more one transformation would be required.

To show the result of tensor rotation on the piezoelectric and elastic constant, the GaAs material system is considered for acoustic waves propagating in the [1,1,0] direction. This direction is used often with this material system as the piezoelectric effect in GaAs does not produce acoustic waves when the exciting electric field is invariant in the [0,1,0] direction (as discussed in 3.2.1). The original constants are

$$\mathbf{e} = \begin{pmatrix} 0 & 0 & 0 & -0.16 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.16 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.16 \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 11.9\times10^{10} & 5.34\times10^{10} & 5.34\times10^{10} & 0 & 0 & 0 \\ 5.34\times10^{10} & 11.9\times10^{10} & 5.34\times10^{10} & 0 & 0 & 0 \\ 5.34\times10^{10} & 5.34\times10^{10} & 11.9\times10^{10} & 0 & 0 & 0 \\ 0 & 0 & 0 & 5.96\times10^{10} & 0 & 0 \\ 0 & 0 & 0 & 0 & 5.96\times10^{10} & 0 \\ 0 & 0 & 0 & 0 & 0 & 5.96\times10^{10}, \end{pmatrix},$$

where $\mathbf{e}$ is the piezoelectric constant and $\mathbf{C}$ is the elastic constant, both in matrix notation, with units of Cm^{-2} and Nm respectively. After a rotation about the x axis of 45°, they become

$$\mathbf{e} = \begin{pmatrix} 0 & 0 & 0 & -0.16 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.16 & 0 \\ 0.08 & -0.08 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 14.58 \times 10^{10} & 2.66 \times 10^{10} & 5.34 \times 10^{10} & 0 & 0 & 0 \\ 2.66 \times 10^{10} & 14.58 \times 10^{10} & 5.34 \times 10^{10} & 0 & 0 & 0 \\ 5.34 \times 10^{10} & 5.34 \times 10^{10} & 11.9 \times 10^{10} & 0 & 0 & 0 \\ 0 & 0 & 0 & 5.96 \times 10^{10} & 0 & 0 \\ 0 & 0 & 0 & 0 & 5.96 \times 10^{10} & 0 \\ 0 & 0 & 0 & 0 & 0 & 3.28 \times 10^{10}, \end{pmatrix},$$

respectively. Clearly, this rotation has a big impact not only on the values of specific elements within the tensors, but also on the positions of the non-zero elements within the tensors, showing that tensor rotation is a very important point to consider when simulating acoustic wave propagation along directions other than the crystal axes.

3.2 Finite-difference time-domain model of acoustic wave propagation

In order to simulate acoustic wave propagation through arbitrary crystalline structures a numerical solution to the acoustic wave equations of motion is required. The finite-difference time-domain method is a powerful tool that allows time-dependent systems of partial-differential equations to be solved simply and efficiently, whilst retaining a generality, that allows a solution to be found for any arbitrary structure, providing the correct boundary conditions are imposed. Furthermore, the concepts of an interlaced mesh as well as the perfectly-matched-layer boundary condition, that have emerged through the continued development of the method, provide a more accurate solution than conventional meshing techniques and ensure that the artificial boundaries imposed at the edge of the simulation domain do not interfere with the solution within the region of interest.

In this section the finite-difference time-domain method is applied the acoustic wave equations of motion to produce a model of acoustic wave propagation in crystalline materials. Firstly, the assumption that the waves under consideration are of Rayleigh-wave type is discussed, which allows the simulation domain to be quasi-three dimensional and greatly reduces the computational effort required compared to a full three dimensional simulation. Then, the finite-difference approximation applied to produce the acoustic-wave model, which of followed by results of acoustic wave propagation within bulk crystalline material as a proof of concept for the model.

3.2.1 Rayleigh-wave assumption of acoustic-wave propagation

Since the ultimate aim of the simulation of acoustic wave propagation is to model surface acoustic wave devices, some consideration of these devices needs to be taken before deriving the model. The striking feature of these devices, from a modelling perspective, is that their geometry, as well as the acoustic waves which they utilise, are practically invariant along the direction parallel to the wave front, compared to the length-scale of the wavelengths they use. A sensible approximation, therefore, is to assume that the geometry is invariant along the direction parallel to the wavefront and the waves have a straight wavefront, i.e. are of Rayleigh-wave type.

Setting the direction parallel to the wavefront to be along the x_2 axis, implementing this approximation then involves setting all terms with a differential with respect to this axis, $\frac{\partial}{\partial x_2}$, equal to zero in the equations governing the acoustic wave propagation. While the differential is set to zero, the displacement along the x_2 may be non-zero and therefore must still be considered within the model. For this reason, using this approximation makes the *quasi*-three dimensional since it is not a full three-dimensional model but contains more information than a two-dimensional model which would assume that all variables along this axes are zero, rather than invariant.

3.2.2 Finite-difference time-domain method

The simplest method of descretising a time-dependent set of equations is to directly apply the finite-difference approximation to the time- and spatial-derivatives. The natural choice of grid for the resulting set of discrete equations then sees every variable calculated at every grid point in space and for every discrete point in space, as depicted in figure 3.3.

While such a discretisation method will work well for many systems of partial differential equations, particularly those containing only second-order differentials, by utilising an interlaced grid the accuracy of the numerical solution may be increased without increasing the number of sampling points required. This may be shown by considering the acoustic wave equations of motion which are to be discretised,

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial \sigma_{ij}}{\partial x_j} \qquad \text{for } i, j = 1, 2, 3, \tag{3.40}$$

where,

$$\sigma_i = C_{ij} \epsilon_j \qquad \text{for } i, j = 1, \ldots, 6. \tag{3.41}$$

By substituting (3.41) into (3.40) then applying the finite difference approximation in both

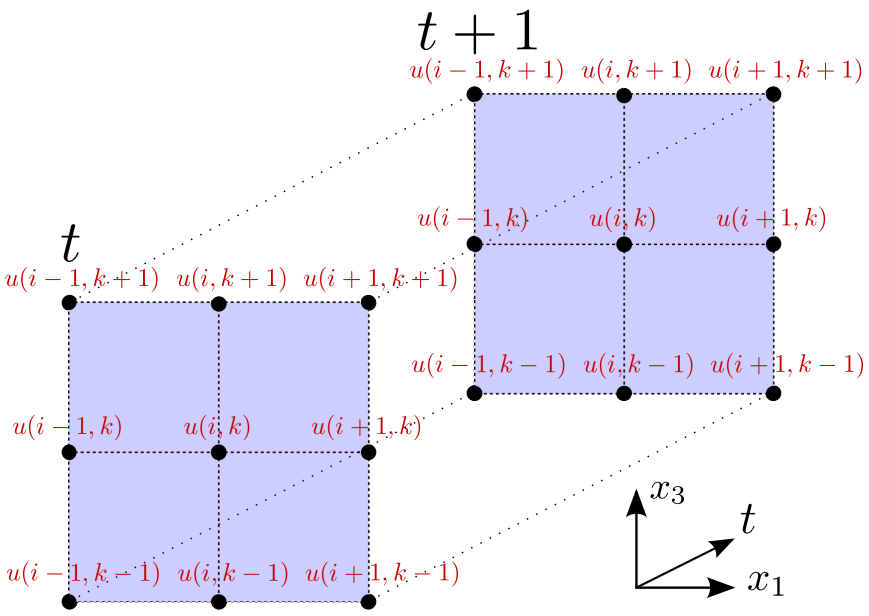

Figure 3.3: The simplest choice of discretisation grid for some spatially and time depedent function, u.

space and time, a set of discrete equations of the form

$$\frac{u_i(i,j,t+1) - 2u_i(i,j,t) + u_i(i,j,t-1)}{(\delta t)^2} = \frac{u_i(i+1,j,t) - 2u_i(i,j,t) + u_i(i-1,j,t)}{(\delta x_1)^2} \ldots \tag{3.42}$$

is obtained, where the other terms have been omitted as they cannont be written succinctly in the Einstein summation convention. Equation (3.42) shows that the spatial differentials will be calculated over a distance of $2\delta z$ (considering only one-dimension), since they depend upon the neighbouring points, and the temporal differentials over a time of $2\delta t$ since the next time step depends on the previous two. If, however, an auxillery variable, $\mathbf{v}$ is added to (3.40) in order to split the time differential, such that

$$\rho \frac{\partial u_i}{\partial t} = \frac{\partial v_{ij}}{\partial x_j} \qquad \text{for } i,j = 1,2,3, \tag{3.43}$$

where,

$$\frac{\partial v_i}{\partial t} = \sigma_i = C_{ij}\epsilon_j \qquad \text{for } i,j = 1,\ldots,6, \tag{3.44}$$

Each equation is now only made up of first-order differentials. Applying the finite-difference

approxiation

$$\frac{u_i(i,j,t+1)-u_i(i,j,t)}{(\delta t)} = \frac{v_i(i+1,j,t)-v_i(i,j,t)}{(\delta x_1)} \ldots,$$
$$\frac{v_i(i,j,t+1)-v_i(i,j,t)}{(\delta t)} = \frac{u_i(i+1,j,t)-u_i(i,j,t)}{(\delta x_1)} \ldots,$$

then gives differentials applied across δz spatially and δt temporally. This alone does not improve the numerical accuracy however, since if these equation where applied to the same discretisation grid as above the same result would be obtained (since the two sets of equations are equivalent). Using an interlaced grid however, where the grid points for the auxiliary variable $\mathbf{v}$ are placed in between the grid points used for the variable $\mathbf{u}$, both spatially and temporally (i.e. at $i \pm \frac{1}{2}$, $j \pm \frac{1}{2}$ and $t \pm \frac{1}{2}$) as depicted in 3.4, each first-order differential is calculated across a distance of $\frac{1}{2}\delta x_i$ or $\frac{1}{2}\delta t$ meaning that the second-order differentials are effectively calculated over δx_i or δt. Therefore, by utilising an interlaced grid the effective accuracy of the calculation can be increased by a factor of 2 without increasing the number of sampling points. The interlaced grid changes (3.45) to

$$\frac{u_i(i,j,t+1)-u_i(i,j,t)}{(\delta t)} = \frac{v_i(i+\frac{1}{2},j,t)-v_i(i-\frac{1}{2},j,t)}{(\delta x_1)} \ldots,$$
$$\frac{v_i(i,j,t+\frac{1}{2})-v_i(i,j,t-\frac{1}{2})}{(\delta t)} = \frac{u_i(i+1,j,t)-u_i(i,j,t)}{(\delta x_1)} \ldots.$$

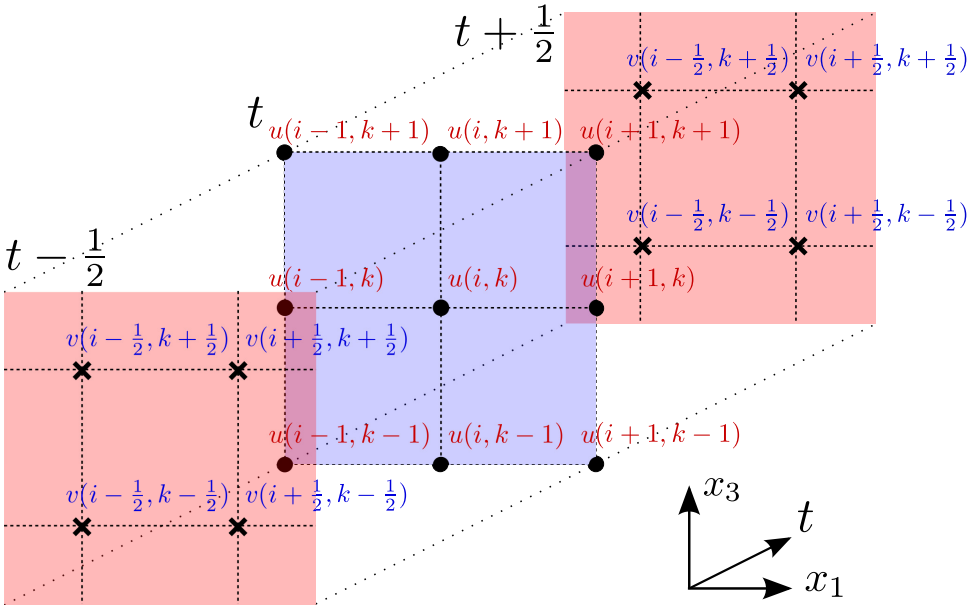

Figure 3.4: Interlaced mesh used in the finite-difference time-domain method.

The disadvantage of using an interlaced grid is the increase in complexity in the implementation of the solution. This comes about as all of the spatial points of one variable at one time step, say $\mathbf{v}|_{t-\frac{1}{2}}$, must be calculated before the next half-time-step of the other variable, i.e. $\mathbf{u}|_t$, meaning that two separate loops over space are required for each whole-time-step, one for each variable. Including the solution of Poisson's equation at each time-step, the flow-diagram for solving the acoustic wave equations of motion using the finite-difference time-domain method with auxiliary variables is shown in figure 3.5.

3.2.3 Solving Poisson's equation in two dimensions

The time evolution of the acoustic-wave equations of motion requires that the potential be found, and therefore Poisson's equation be solved at every time-step. Furthermore, since the simulation domain considered is quasi-three dimensional and it is only the electric field (i.e. $-\frac{\partial\phi}{\partial x_i}$) that appears within the equations of motion, Poisson's equation only needs to be solved in two-dimensions as one component of electric field will always be zero. This is where the biggest computational *saving* is made from the quasi-three dimensional approximation as solving Poisson's equation in three dimensions using the matrix diagonalisation approach described below makes the runtime of this method scale roughly cubicly with the number of sampling points, meaning that when simulating larger devices of a real-world size, solving Poisson's equation would take up nearly all of the total runtime of the FDTD algorithm.

The notation used within this subsection, particularly with reference to axis labels, has been made general (as opposed to the matrix-style axis labels used when deriving the acoustic-wave equations of motion) because the solution presented here is a general solution to the two-dimensional Poisson equation. Furthermore, the notation has been changed to emphasise the point that the axes used here are chosen arbitrarily and can be aligned along any of the axes defined in the acoustic-wave equations of motion. In fact, the orientation of this solution to Poisson's equation is actually a very important consideration to the efficiency of the solution, as discussed later in this subsection.

Derivation

The general form of Poisson's equation is

$$\nabla \cdot (\varepsilon \nabla \phi) = -\rho, \tag{3.45}$$

where ε is the effective permittivity, ϕ is the potential and ρ is the charge density. Since the Poisson equation is to be discretised over two-dimensions the ∇ operator is expanded over

two orthogonal axes x and y to give

$$\frac{\partial}{\partial x}\varepsilon\left(\frac{\partial}{\partial x}\phi\right)+\frac{\partial}{\partial y}\varepsilon\left(\frac{\partial}{\partial y}\phi\right)=-\rho. \tag{3.46}$$

Applying the finite difference approximation gives

$$\begin{gathered}\left\{\frac{\varepsilon_{i+1,j}\left(\frac{\phi_{i+1,j}-\phi_{i-1,j}}{2(\delta x)}\right)\Big|_{i+1,j}-\varepsilon_{i-1,j}\left(\frac{\phi_{i+1,j}-\phi_{i-1,j}}{2(\delta x)}\right)\Big|_{i-1,j}}{2(\delta x)}\right\}+\\ \left\{\frac{\varepsilon_{i,j+1}\left(\frac{\phi_{i,j+1}-\phi_{i,j-1}}{2(\delta x)}\right)\Big|_{i,j+1}-\varepsilon_{i,j-1}\left(\frac{\phi_{i,j+1}-\phi_{i,j-1}}{2(\delta x)}\right)\Big|_{i,j-1}}{2(\delta y)}\right\}=-\rho,\end{gathered} \tag{3.47}$$

where δx and δy and the spatial discretisation steps along the x and y axes respectively. Rearranging then gives

$$\begin{gathered}\left\{\frac{\varepsilon_{i+1,j}\left(\phi_{i+2,j}-\phi_{i,j}\right)-\varepsilon_{i-1,j}\left(\phi_{i,j}-\phi_{i-2,j}\right)}{4(\delta x)^2}\right\}\\ \left\{\frac{\varepsilon_{i,j+1}\left(\phi_{i,j+2}-\phi_{i,j}\right)-\varepsilon_{i,j-1}\left(\phi_{i,j}-\phi_{i,j-2}\right)}{4(\delta y)^2}\right\}=-\rho.\end{gathered} \tag{3.48}$$

Making the substitutions $2(\delta x)\rightarrow\delta x$ and $2(\delta y)\rightarrow\delta y$, as in appendix A, gives

$$\begin{gathered}\left\{\frac{\varepsilon_{i+\frac{1}{2},j}}{(\delta x)^2}\phi_{i+1,j}+\frac{\varepsilon_{i-\frac{1}{2},j}}{(\delta x)^2}\phi_{i-1,j}+\frac{\varepsilon_{i,j+\frac{1}{2}}}{(\delta y)^2}\phi_{i,j+1}+\frac{\varepsilon_{i,j-\frac{1}{2}}}{(\delta y)^2}\phi_{i,j-1}-\right.\\ \left.\left(\frac{\varepsilon_{i+\frac{1}{2},j}+\varepsilon_{i-\frac{1}{2},j}}{(\delta x)^2}+\frac{\varepsilon_{i,j+\frac{1}{2}}+\varepsilon_{i,j-\frac{1}{2}}}{(\delta y)^2}\right)\phi_{i,j}\right\}=-\rho\end{gathered} \tag{3.49}$$

where variables indexed by $\pm\frac{1}{2}$ are found by averaging.

The system of linear equations that eq (3.49) represents may be arranged into a matrix equation of the form

$$\mathbf{A}\phi=-\rho, \tag{3.50}$$

where ϕ and ρ are vectors of all the spatial points of each variable arranged such that each row of spatial points is placed one after another in ascending row order, i.e.

$$\phi=\left[\phi_{1,1},\phi_{2,1},\phi_{3,1},\cdots,\phi_{ni-1,1},\phi_{ni,1},\phi_{1,2},\phi_{2,2},\cdots,\phi_{ni-1,nj},\phi_{ni,nj}\right],$$

$$\rho=\left[\rho_{1,1},\rho_{2,1},\rho_{3,1},\cdots,\rho_{ni-1,1},\rho_{ni,1},\rho_{1,2},\rho_{2,2},\cdots,\rho_{ni-1,nj},\rho_{ni,nj}\right],$$

and $\mathbf{A}$ is a symmetric, sparse-banded matrix of the order of $ninj$, given by

$$\mathbf{A}=\begin{bmatrix}\mathbf{B_1} & \frac{\varepsilon_{i,\frac{1}{2}}}{(\delta y)^2}\mathbf{I} & & & & \\ \frac{\varepsilon_{i,\frac{1}{2}}}{(\delta y)^2}\mathbf{I} & \mathbf{B_2} & \frac{\varepsilon_{i,1\frac{1}{2}}}{(\delta y)^2}\mathbf{I} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \frac{\varepsilon_{i,nj-1\frac{1}{2}}}{(\delta y)^2}\mathbf{I} & \mathbf{B_{nj-1}} & \frac{\varepsilon_{i,nj-\frac{1}{2}}}{(\delta y)^2}\mathbf{I} & \\ & & & \frac{\varepsilon_{i,nj-\frac{1}{2}}}{(\delta y)^2}\mathbf{I} & \mathbf{B_{nj}} & \end{bmatrix},$$

where $\mathbf{I}$ is the identity matrix of the order of ni and $\mathbf{B}$ is a symmetric, tridiagonal matrix given by

$$\mathbf{B_j} = \begin{bmatrix} a_{1,j} & b_{2,j} & & & & \\ b_{2,j} & a_{2,j} & b_{3,j} & & & \\ & & \ddots & \ddots & \ddots & \\ & & & & b_{ni,j} & a_{ni,j} \end{bmatrix}$$

and a and b are

$$a_{i,j} = \left(\frac{\varepsilon_{i+\frac{1}{2},j} + \varepsilon_{i-\frac{1}{2},j}}{(\delta x)^2} + \frac{\varepsilon_{i,j+\frac{1}{2}} + \varepsilon_{i,j-\frac{1}{2}}}{(\delta y)^2} \right),$$

$$b_{i,j} = \frac{\varepsilon_{i-\frac{1}{2},j}}{(\delta x)^2}.$$

Note that in the construction of $\mathbf{A}$, Neumann boundary conditions have been used. While other boundary conditions could be added, such as those described in 2.2.1, this is not required as the current system under consideration is always assumed to be isolated from external electric fields.

Solution

The equation 3.50 is solved by passing the matrix to diagonalise, $\mathbf{A}$, along with the charge distribution, ρ, to the LAPACK linear algebra package [47] which returns the solution.

As a test case, Poisson's equation was solved for a square of bulk GaAs, $30 \times 30\,\mu$m in size. A charge of 1 Coulomb was placed on each of the 9 grid points in the 3×3 square at the centre of the simulation domain, giving a charge density of approximately $44.44 \times 10^{12}\,\mathrm{Cm}^{-2}$. The discretisation steps, δx and δy, were both set to $0.15\,\mu$m. Figure 3.6 shows the solution.

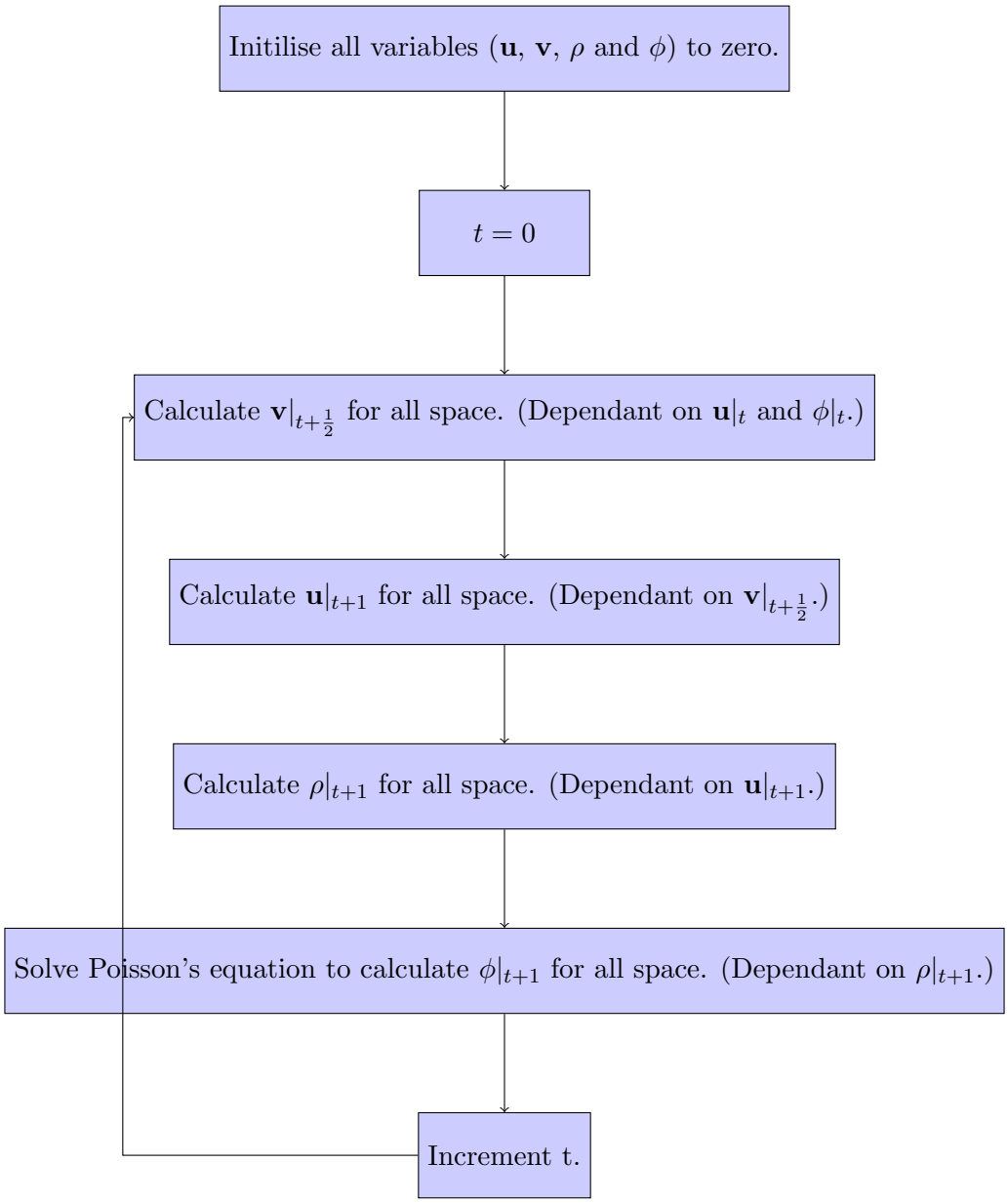

Figure 3.5: Flow diagram of the finite-difference time-domain solution of the acoustic wave equations of motion for the displacement, **u**, the auxiliary variable, **v**, the induced charged displacement, ρ, and the potential, ϕ.

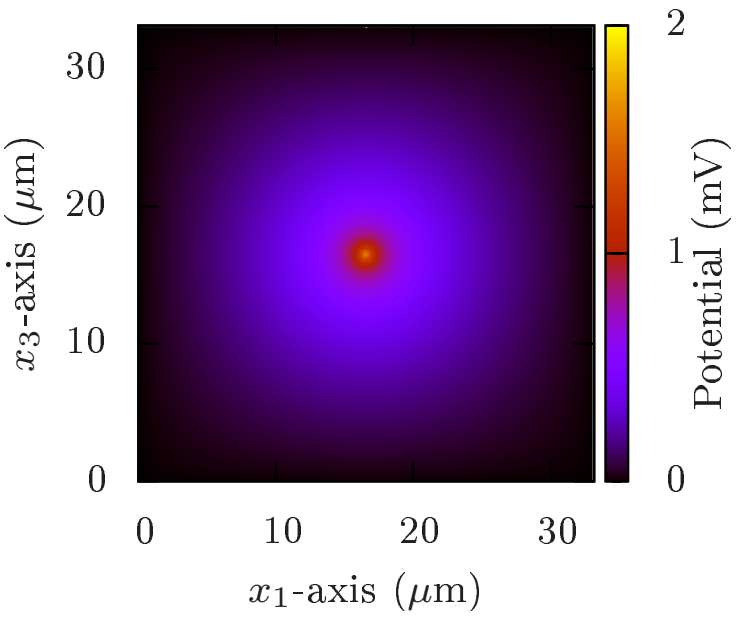

Figure 3.6: Solution of Poisson's equation in two dimensions within bulk GaAs with a $0.45 \times 0.45\,\mu$m square of charge at the centre of the simulation and a charge density of $44.44\,\mathrm{Cm}^{-2}$.

Computational efficiency of the solution

Since the matrix to be solved is of a symmetric, banded form, the most efficient method for solving the matrix is to use a symmetric-banded matrix solver such as that provided by LAPACK [47] which ignores the zero-elements outside the outer-most diagonal, therefore only diagonalising the inner-most, non-zero part of the matrix. This reduces the number of operations required from the order of $(ninj)^2$ for Gaussian elimination, down to the order of $ni(nj)^2$. From this, it is clear to see that the dimension nj has the biggest effect on the number of opperations required. Furthermore, since the axis in this solution to Poisson's equation are not specific to either axis within the acoustic wave simulation, then the y axis of Poisson's equation may be aligned to the shorter axis of the of the acoustic wave simulation domain (if it is not square), hence making the solution more efficient. Clearly, this effect is more pronounced on simulation domains where the ratio of the two sides are far from 1. When simulating surface acoustic wave devices in particular, as in 4, where one side of the simulation domain is often more than several orders of magnitude bigger than the other, it is imperative that this effect be considered.

The second point on efficiency which should be considered when solving Poisson's equation within an acoustic wave simulation, is that Poisson's equation must be solved at every time step. Furthermore, the form of the matrix to solve does not change over time, only the charge density varies, therefore there is no need to diagonalise the matrix at every time step. Instead, if the diagonalised matrix is stored then this may be used to calculate the new potential at every time step. LAPACK [47] offers this functionality by using its diver routines, whereby the matrix may be diagonalised before the loop over time is entered and then new potentials found by supplying a new charge density at every time step. Since the majority of computational effort is spent diagonalising the matrix when solving Poisson's equation, this can have a dramatic improvement in the runtime of a simulation, especially when there are many iterations over time.

3.2.4 Solution of acoustic wave equations of motion

The acoustic wave equations of motion are solved for a square of bulk GaAs in order to demonstrate the viability of the finite-difference time-domain method. Since GaAs does not produce propagating acoustic waves when an electric field is applied parallel to the crystal axes, the axes of the simulation domain are rotated such that the x-axis lies along the [1,1,0] crystal direction. This involves the rotation of the piezoelectric and elastic tensor constants as

described in 3.1.5. The system is excited by placing a square block of charge in the centre of the simulation domain, as in 3.2.3, the magnitude of which sinusoidally oscillates at frequency of 1 GHz. This generates a sinusoidally varying potential which in term launches an acoustic wave. The speed of sound within GaAs is approximately $2800\,\mathrm{ms}^{-1}$ giving a wavelength of $2.8\,\mu$m. The discretisation steps within the simulation, δx_1 and δx_3, were therefore set to $0.15\,\mu$m. Figure 3.7 shows various time frames of the solution for displacement in the x_1 and x_3 directions, u_1 and u_3 respectively. The displacement in the x_2 direction, u_2, is always zero within GaAs in the [1,1,0] orientation. This clearly shows the acoustic wave that is generate which propagates radially outward as expected.

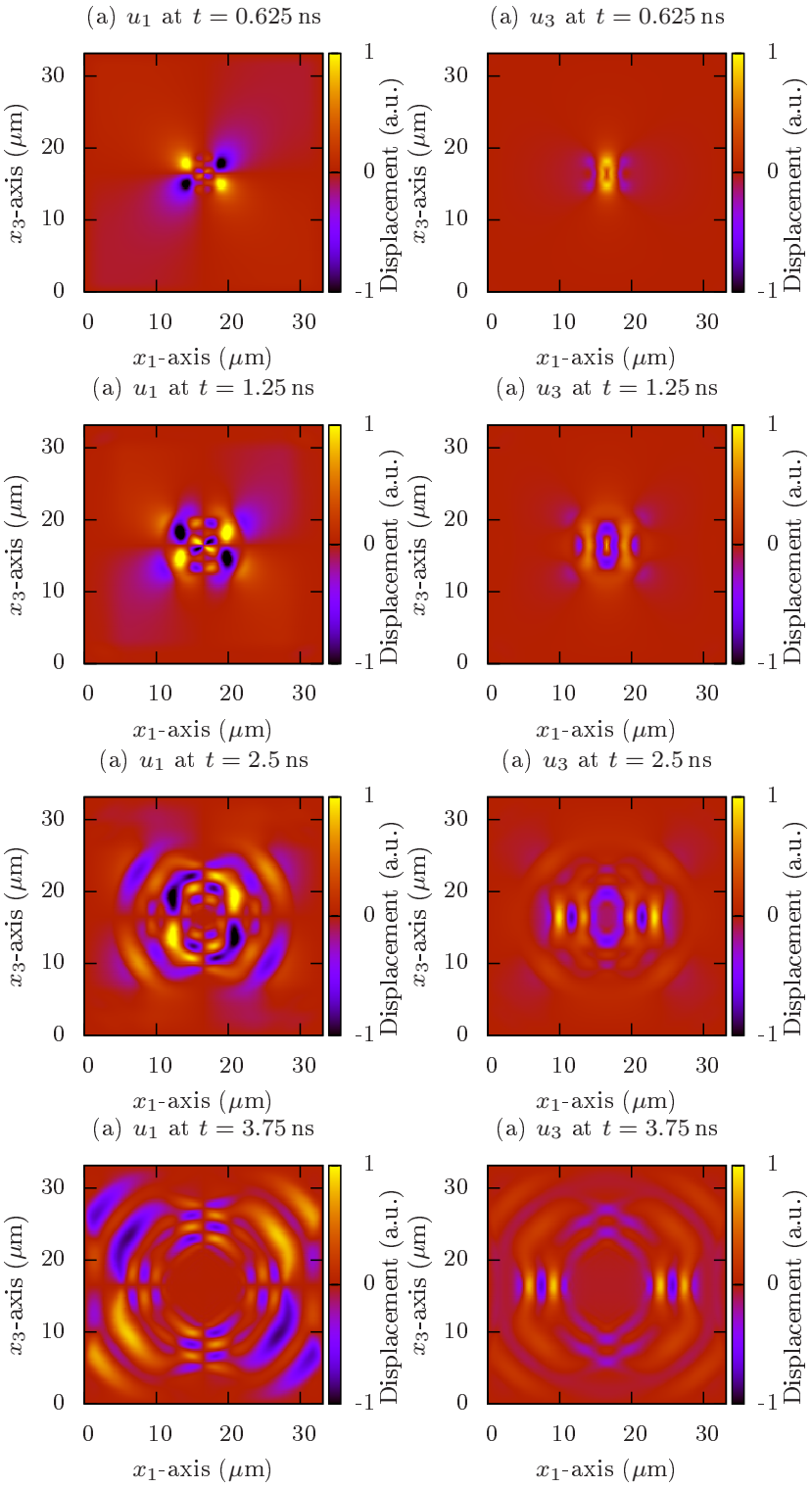

Figure 3.7: Solution of the acoustic wave equations of motion in the [1,1,0] orientated, GaAs material system. The acoustic wave excited by sinusoidally oscillating the magnitude of a block of charge in the central $0.45 \times 0.45\,\mu$m of the simulation domain (i.e. a 3×3 square of grid points). The charge was oscillated at a frequency of 1 GHz producing for an acoustic wave which propagates radially outwards from the centre of the simulation domain.

3.3 Perfectly-matched-layer boundary conditions

The perfectly-matched-layer (PML), first introduced in 1994 [52], is a powerful numerical technique which allows travelling waves within free-space simulations to be completely absorbed. It is used as a boundary condition within simulations of wave propagation in order to stop artificial reflections from the boundaries of the simulation domain interfering with the region of interest. The success of the PML over other types of absorbing boundary condition (ABC) comes from the fact that, before discretisation, the equations within the boundary layer are perfectly matched to those within the region of interest, hence the name. This allows for the absorption coefficient to be changed very rapidly, not only between the region of interest and PML but within the PML itself, without causing instabilities. Any wave entering the PML is therefore damped very effectively with PMLs whose thickness is just a fraction of a wavelength will reduced the amplitude of the reflected wave by several orders of magnitude.

PML were initially generated for electromagnetic problems by numerically calculating the parameters of the PML in order to give an (almost) reflectionless absorber.[52] Since then, the method has been mathematically proven in many different ways which see the PML, broadly, as either an artificial anisotropic absorbing medium [53] or an analytic continuation of spatial variables onto the complex plane [54]. In this thesis, that latter method is considered as its application is better suited for large systems of PDEs.

While the concept of PMLs has been applied to domains of acoustic waves before in the form of elastodynamics [55], they have never been applied to acoustic wave propagation in piezoelectric crystals before the current work. PMLs for acoustic wave propagation in piezoelectric crystals are derived in this section.

3.3.1 Analytic continuation of spatial variables onto the complex plane

The action of a PML may be thought of as an analytic continuation of spatial variables onto the complex plane. Therefore as a travelling wave enters the PML, it will no longer travel along purely real spatial coordinates, but instead will have some non-zero complex component of spatial coordinates also. In order to see what effect this has, consider a plane wave in one dimension

$$w(x,t) = Ae^{i(kx-\omega t)}, \tag{3.51}$$

with amplitude A, wave vector k and angular frequency ω. Once the wave enters the PML, the spatial variable, x, becomes complex and therefore (3.51) becomes

$$\begin{aligned} w(x,t) &= Ae^{i(k(\Re x + i\Im x) - \omega t)} \\ &= Ae^{-k\Im x} e^{i(k\Re x - \omega t)}, \end{aligned} \tag{3.52}$$

such that wave exponentially decays as it travels through the PML, hence making the PML act as an absorbing material, as shown in 3.8. The key difference between a PML and a regular absorbing boundary is that the plane wave solution has not changed after entering the PML, only the coordinate space onto which it is mapped has changed. This ensures that any wave entering the PML will be unaffected by this absorption and will not reflect back, therefore making the layer *perfectly matched* and reflectionless.

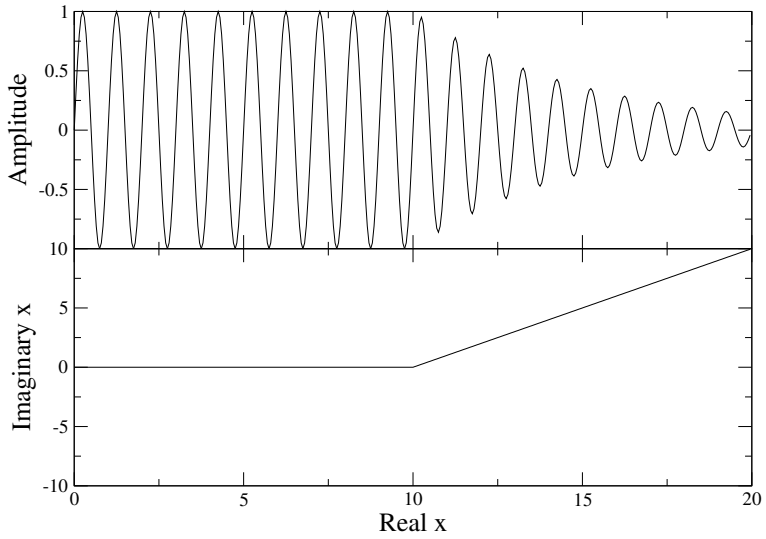

Figure 3.8: Concept of a PML. Within the PML (the area > 10), the x coordinate gains an miagainary component causing the wave to exponentially decay.

In order to make the implementation of the PML simpler, the complex space within the PML may be transformed back onto the real axis. This transformation requires a substitution of variables, the first stage of which is to define the complex space in terms of the real space it is be transformed to, say,

$$\mathbf{x}(\Re x) = \Re x + if(\Re x), \tag{3.53}$$

where $\mathbf{x}$ is complex space and $f(\Re x)$ is some function defining the magnitude of the complex component. Because, the PML is to be applied within a system of partial differential equations, (3.53) is differentiated with respect to x to give a differential relation,

$$\partial \mathbf{x} = (1 + i\frac{\partial f(x)}{\partial x})\partial(\Re x). \tag{3.54}$$

The form of the function f, or more precisely $\frac{\partial f(x)}{\partial x}$ since this is what will be used deriving the PML boundary conditions, is now sought as this will determine the attenuation within the PML region. Defining

$$\frac{\partial f(x)}{\partial x} = \zeta(x)$$

where ζ determines the magnitude of the complex component, and performing the substitution of variables in (3.51) gives

$$w(x,t) = Ae^{-k\int_x \zeta(x)x} e^{i(k\Re x - \omega t)}.$$

where the attenuation will be dependent upon wavelength, with longer wavelengths attenuated more that shorter ones because of the dependence on wave vector in the exponential decay term. However, if f is given the form

$$\frac{\partial f(x)}{\partial x} = \frac{\zeta(x)}{\omega},$$

then performing the substitution of variables in (3.51) gives

$$w(x,t) = Ae^{-\frac{k}{\omega}\int_x \zeta(x)x} e^{i(k\Re x - \omega t)}.$$

where the attenuation is independent of the wavelength, assuming the material is dispersionless, since $\frac{k}{\omega} = \frac{1}{c_{ph}}$ is the inverse of the phase velocity which is constant for all wavelengths in a dispersionless material. This clearly shows that the form of f plays an integral role in the behaviour of PMLs, although a full discussion of the implications of the form of f are beyond the scope of this thesis and more details can be found in [56]. In this thesis, the form

$$\frac{\partial f(x)}{\partial x} = \frac{\zeta(x)}{\omega}, \tag{3.55}$$

is used because it is independent of frequency, a property which is very useful when simulating in the time domain since the frequencies generated within the simulation domain are not calculated. This gives the relation for the substitution of variables, in terms of differentials, used to derive the PML boundary conditions, as

$$\frac{\partial}{\partial x} = \frac{1}{1 + i\frac{\zeta(x)}{\omega}} \frac{\partial}{\partial x}. \tag{3.56}$$

3.3.2 Substitution of variables

In order to implement PML boundary conditions a substitution of variables must be performed on the acoustic wave equations of motion

$$\rho \frac{\partial u_i}{\partial t} = \frac{\partial v_{ij}}{\partial x_j}, \tag{3.57}$$

$$\frac{\partial v_i}{\partial t} = \sigma_i = C_{ij}\epsilon_j + e_{ik}^{\mathrm{T}}\frac{\partial \phi}{\partial x_k}. \tag{3.58}$$

This is done by transforming to the frequency domain, such that $\mathbf{u}(t) \rightarrow \mathbf{U}(\omega)$ and $\mathbf{v}(t) \rightarrow \mathbf{V}(\omega)$, we obtain

$$-\mathrm{i}\omega\rho U_i = \frac{\partial V_{ij}}{\partial x_j}, \tag{3.59}$$

and

$$-\mathrm{i}\omega V_i = C_{ij}\epsilon_j + e_{ik}^{\mathrm{T}}\frac{\partial \phi}{\partial x_k}, \tag{3.60}$$

respectively, where ω is the angular frequency of the acoustic wave. A complex change of variables is now introduced such that $x_i \rightarrow (1+\mathrm{i}\frac{\zeta_i}{\omega})x_i$, where ζ has two components, $\zeta_1(x_1)$ and $\zeta_3(x_3)$, which are the absorption profiles of the PML in the x_1 and x_3 direction respectively, to give

$$-\mathrm{i}\omega\rho U_i = \frac{1}{1+\mathrm{i}\frac{\zeta_j}{\omega}}\frac{\partial V_{ij}}{\partial x_j}, \tag{3.61}$$

and

$$\begin{aligned} -\mathrm{i}\omega V_i = {} & C_{i1}\frac{1}{1+\mathrm{i}\frac{\zeta_1}{\omega}}\frac{\partial U_1}{\partial x_1} + C_{i3}\frac{1}{1+\mathrm{i}\frac{\zeta_3}{\omega}}\frac{\partial U_3}{\partial x_3} + C_{i4}\frac{1}{1+\mathrm{i}\frac{\zeta_3}{\omega}}\frac{\partial U_2}{\partial x_3} \\ & + C_{i5}\left(\frac{1}{1+\mathrm{i}\frac{\zeta_3}{\omega}}\frac{\partial U_1}{\partial x_3} + \frac{1}{1+\mathrm{i}\frac{\zeta_1}{\omega}}\frac{\partial U_3}{\partial x_1}\right) + C_{i6}\frac{1}{1+\mathrm{i}\frac{\zeta_1}{\omega}}\frac{\partial U_2}{\partial x_1} \\ & + e_{i1}^{\mathrm{T}}\frac{1}{1+\mathrm{i}\frac{\zeta_1}{\omega}}\frac{\partial \phi}{\partial x_1} + e_{i3}^{\mathrm{T}}\frac{1}{1+\mathrm{i}\frac{\zeta_3}{\omega}}\frac{\partial \phi}{\partial x_3}. \end{aligned} \tag{3.62}$$

Multiplying by $(1+\mathrm{i}\frac{\zeta_1}{\omega})(1+\mathrm{i}\frac{\zeta_3}{\omega})$ gives

$$-\mathrm{i}\omega\rho U_i + \rho(\zeta_1+\zeta_3)U_i + \frac{\mathrm{i}}{\omega}\rho\zeta_1\zeta_3 U_i = \frac{\partial V_{ij}}{\partial x_j} + \frac{\mathrm{i}}{\omega}\left(\zeta_1\frac{\partial V_{i3}}{\partial x_3} + \zeta_3\frac{\partial V_{i1}}{\partial x_1}\right), \tag{3.63}$$

and

$$\begin{aligned} -\mathrm{i}\omega V_i + (\zeta_1+\zeta_3)V_i + \frac{\mathrm{i}}{\omega}\zeta_1\zeta_3 V_i = {} & C_{ij}\epsilon_j + e_{ik}^{\mathrm{T}}\frac{\partial \phi}{\partial x_k} + \\ & \frac{\mathrm{i}}{\omega}\left(\zeta_1\left\{C_{i3}\frac{\partial U_3}{\partial x_3} + C_{i4}\frac{\partial U_2}{\partial x_3} + C_{i5}\frac{\partial U_1}{\partial x_3} + e_{i3}^{\mathrm{T}}\frac{\partial \phi}{\partial x_3}\right\}\right. \\ & \left. + \zeta_3\left\{C_{i1}\frac{\partial U_1}{\partial x_1} + C_{i5}\frac{\partial U_3}{\partial x_1} + C_{i6}\frac{\partial U_2}{\partial x_1} + e_{i1}^{\mathrm{T}}\frac{\partial \phi}{\partial x_1}\right\}\right). \end{aligned} \tag{3.64}$$

Transforming back to the time domain gives

$$\rho\frac{\partial u_i}{\partial t} = \frac{\partial v_{ij}}{\partial x_j} - \rho(\zeta_1+\zeta_3)u_i + \alpha_i, \tag{3.65}$$

where the auxiliary field α has been introduced in place of the $\frac{\mathrm{i}}{\omega}$ terms in (3.63), which become time integrals when transformed to the time domain, such that the time derivative of α is

$$\frac{\partial \alpha_i}{\partial t} = \zeta_1\frac{\partial v_{i3}}{\partial x_3} + \zeta_3\frac{\partial v_{i1}}{\partial x_1} - \rho\zeta_1\zeta_3 u_i, \tag{3.66}$$

and

$$\frac{\partial v_i}{\partial t} = C_{ij}\epsilon_j + e^{\mathrm{T}}_{ik}\frac{\partial \phi}{\partial x_k} - (\zeta_1 + \zeta_3)v_i + \beta_i, \tag{3.67}$$

where the auxiliary field β has been introduced in place of the $\frac{\mathrm{i}}{\omega}$ terms in (3.64), such that its time derivative is

$$\begin{aligned}\frac{\partial \beta_i}{\partial t} = \zeta_1 &\left\{C_{i3}\frac{\partial u_3}{\partial x_3} + C_{i4}\frac{\partial u_2}{\partial x_3} + C_{i5}\frac{\partial u_1}{\partial x_3} + e^{\mathrm{T}}_{i3}\frac{\partial \phi}{\partial x_3}\right\} \\ + \zeta_3 &\left\{C_{i1}\frac{\partial u_1}{\partial x_1} + C_{i5}\frac{\partial u_3}{\partial x_1} + C_{i6}\frac{\partial u_2}{\partial x_1} + e^{\mathrm{T}}_{i1}\frac{\partial \phi}{\partial x_1}\right\} - \zeta_1\zeta_3 v_i.\end{aligned} \tag{3.68}$$

3.3.3 Implementation

Equations (3.65), (3.66), (3.67) and (3.68) represent a new set of equations of motion which will alter the way acoustic waves propagate with the PMLs (namely making them exponentially decay in amplitude). These equations may be implemented in the same manner as the unmodified equations of motion, as described in 3.2.2, however there are some nuances to the implementation which are described below.

As identified by Smith *et al.* [57], the most natural choice of discretisation grid for applying FDTD analysis to the acoustic wave equations of motion is an interlaced mesh, with grid points interlaced in both space and time, as discussed in 3.2.2. Here, u and ϕ are mapped to whole-integer values of x_1, x_3 and t, while v is mapped to half-integer values. Applying the finite difference approximation, it then follows that a spatial derivative depends upon the field values at $\pm\frac{1}{2}$, which lie in between grid points. These *midpoints* may be taken as the average of their adjacent points, i.e. for a spatial derivative in the direction x_1 at the point (i, k)

$$\frac{\partial v}{\partial x_1}(i,k) = \frac{\bar{v}(i+\frac{1}{2},k) - \bar{v}(i-\frac{1}{2},k)}{\delta x} \tag{3.69}$$

where $\bar{v}$ represents the average of the neighbouring points, i.e.

$$\bar{v}(i+\tfrac{1}{2},k) = \frac{v(i+\frac{1}{2},k+\frac{1}{2}) + v(i+\frac{1}{2},k-\frac{1}{2})}{2} \tag{3.70}$$

As is normal with FDTD analysis, the instantaneous values of time-dependent variables are sampled midway between the time steps used in evaluating time-derivatives, such that (3.65) and (3.67) become

$$\rho\frac{u_i|_{t+1} - u_i|_t}{\delta t} = \left[\frac{\partial v_{ij}}{\partial x_j} - \rho(\zeta_1 + \zeta_3)u_i + \alpha_i\right]_{t+\frac{1}{2}}, \tag{3.71}$$

and

$$\frac{v_i|_{t+\frac{1}{2}} - v_i|_{t-\frac{1}{2}}}{\delta t} = \left[C_{ij}\epsilon_j + e^{\mathrm{T}}_{ik}\frac{\partial \phi}{\partial x_k} - (\zeta_1 + \zeta_3)v_i + \beta_i\right]_t. \tag{3.72}$$

However, this implies a dependence on u and v at time steps between those to which these variables are mapped (i.e. $u_i|_{t+\frac{1}{2}}$ in (3.71) and $v_i|_t$ in (3.72)). This may be overcome by first noting that these terms are zero within the region of interest and therefore any approximation that is made will not affect what we are trying to observe except possibly increase artificial reflections from the boundaries, and second realising that the difference between the spatial derivative at half time steps should be small provided that a small time step, δt is used. We therefore make the assumption that, in (3.71), $u_i|_{t+\frac{1}{2}} \approx u_i|_t$, and in (3.72), $v_i|_t \approx v_i|_{t-\frac{1}{2}}$ and note that doing so does not produce a noticeable increase in numerical noise from the PMLs in practice.

A similar problem arises when discretising the auxiliary fields α and β in time

$$\frac{\alpha_i|_{t+\frac{1}{2}} - \alpha_i|_{t-\frac{1}{2}}}{\delta t} = \left[\zeta_1 \frac{\partial v_{i3}}{\partial x_3} + \zeta_3 \frac{\partial v_{i3}}{\partial x_1} - \rho\zeta_1\zeta_3 u_i\right]_t, \tag{3.73}$$

$$\begin{aligned}\frac{\beta_i|_t - \beta_i|_{t-1}}{\delta t} = \Bigg[&\zeta_1 \left\{C_{i3}\frac{\partial u_3}{\partial x_3} + C_{i4}\frac{\partial u_2}{\partial x_3} + C_{i5}\frac{\partial u_1}{\partial x_3} + e^{\mathrm{T}}_{i3}\frac{\partial \phi}{\partial x_3}\right\} \\ &+\zeta_3 \left\{C_{i1}\frac{\partial u_1}{\partial x_1} + C_{i5}\frac{\partial u_3}{\partial x_1} + C_{i6}\frac{\partial u_2}{\partial x_1} + e^{\mathrm{T}}_{i1}\frac{\partial \phi}{\partial x_1}\right\} - \zeta_1\zeta_3 v_i\Bigg]_{t-\frac{1}{2}}.\end{aligned} \tag{3.74}$$

However, here the values of $\mathbf{u}$ and $\mathbf{v}$ at the half time step may be taken as the average of the adjacent time steps since they are already known, i.e.

$$u_i|_{t-\frac{1}{2}} = \frac{u_i|_t + u_i|_{t-1}}{2}, \tag{3.75}$$

and

$$v_{ij}|_t = \frac{v_{ij}|_{t+\frac{1}{2}} + v_{ij}|_{t-\frac{1}{2}}}{2}, \tag{3.76}$$

although we note, however, for simplicity, the same approximation as in (3.71) and (3.72) may be used and in practice the PMLs still have the desired effect.

The simulation domain for time-dependent variables is terminated by hard-wall boundaries so that the simulation takes the form shown in Fig. 3.9. For ϕ, Neumann boundary conditions are implemented. The hard-wall boundary condition is implemented on the whole-integer grid points, mapped to the variable $\mathbf{u}$, meaning that the outside edge of the PML, and therefore the maximum value of ζ, occurs at the half-integer grid points, mapped to $\mathbf{v}$, immediately inside of the hard-wall boundary.

3.3.4 Numerical results

Prior to the current application of PMLs to the acoustic wave equations of motion in piezoelectric crystals, the best method of stopping artificial reflection off the simulation boundaries

$\zeta_x \neq 0$ $\zeta_z \neq 0$	$\zeta_z \neq 0$	$\zeta_x \neq 0$ $\zeta_z \neq 0$
$\zeta_x \neq 0$	Region of interest $\zeta_x = 0, \zeta_z = 0$	$\zeta_x \neq 0$
$\zeta_x \neq 0$ $\zeta_z \neq 0$	$\zeta_z \neq 0$	$\zeta_x \neq 0$ $\zeta_z \neq 0$

Figure 3.9: Illustration of the simulation domain showing the PMLs at the edges of the region of interest and where ζ is non-zero. Note that because ζ is zero within the region of interest the auxiliary fields α and β are also zero here.

was to add a damping coefficient into the equations of motion, as presented by Smith *et al.* [57]. While this method was stable for many material systems, it proved to produce instabilities within the bismuth germanate ($Bi_4Ge_3O_{12}$) material system. In order to demonstrate the effectiveness of the current method, as well as to show its stability, it is applied to the bismuth germanate.

The x_1- and x_3-axes of the simulation are aligned along the [1,1,0] and [0,0,1] crystal axes respectively by rotating the elastic and piezoelectric tensors by 45° about the [0,0,1] crystal-axis such that an acoustic wave will propagate along the [1,1,0] direction, as is done experimentally with cubic crystals. The excitation frequency used was 1 GHz, as in [58], making the spatial discretisation step, δx_1 and δx_3, 0.15 μm (i.e., $\sim \frac{\lambda}{20}$). The grid size of the region of interest was set to 201×201 points and the time step, δt was set to 25 ps observing the FDTD stability criterion [59]. The PMLs used were 10-points thick ($\sim \frac{\lambda}{2}$) and the parameter ζ was increased quadratically up to its maximum stable value of $1.58 \times 10^{10}\,\mathrm{rad\,s^{-1}}$ as described in the following section.

The acoustic wave was excited by setting a 3×3 square in the centre of the simulation domain to have a constant charge which was then oscillated sinusoidally at the excitation frequency for two periods and solving Poisson's equation to find the potential profile around this charge over the simulation domain. This potential profile was used as an input to the acoustic wave equations of motion to launch a propagating wave. This method of excitation was used firstly because it is more physically realistic than exciting a component of the displacement since acoustic waves in piezoelectric crystals are generated using an alternating potential, and secondly because excitations at a single point tend to lead to instabilities

caused by the change in sign of the spatial differential from the point of excitation to its surrounding points. These instabilities manifest themselves as changes in the sign of the solution from one grid point to the next such that the entire solution appears modulated by a sawtooth wave with oscillations of the order of the grid spacing. Fig. 3.10 shows the resulting acoustic wave propagating in bulk bismuth germanate, with material parameters taken from [60], which is absorbed by the PML boundaries.

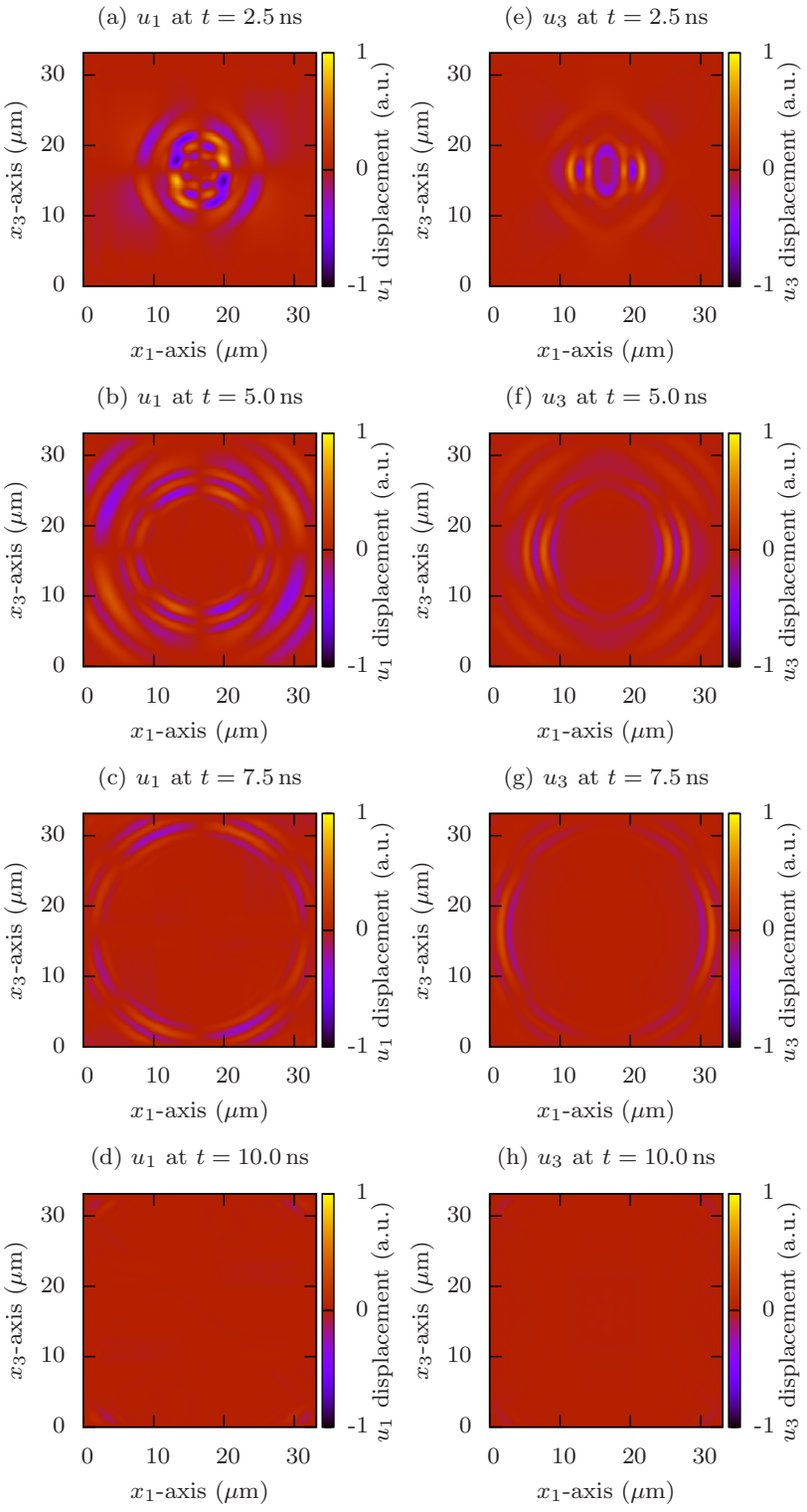

Figure 3.10: The u_1 and u_3 displacement components for the acoustic wave launched in $Bi_4Ge_3O_{12}$ showing how the wave propagates radially outwards and is absorbed by the PML. (Animations of the propagating wave are included as supplementary material. Animation 1 for the u_1 displacement and animation 2 for the u_3 displacement.)

3.3.5 Stability and optimisation of the PML

The discretized PML equations will be subject to a system-dependent stability criterion, much the same as the stability criterion for the unmodified equations within the region of interest. Considering the stability criterion for FDTD analysis in two dimensions [59]

$$v_{\max}\delta t = \left(\frac{1}{\delta x_1^2} + \frac{1}{\delta x_3^2}\right)^{-\frac{1}{2}}, \tag{3.77}$$

where δt, δx_1 and δx_3 are the discretisation parameters in time and space respectively and $v_{\max}$ is the maximum velocity within the simulation domain, it is clear that within the PML layers extra terms will be added to this criterion which have a dependence upon ζ (or more accurately a dependence upon the maximum value of ζ within the PML since this represents the *worst case* scenario). It is important for the maximum stable value of ζ to be found since the attenuation of a propagating wave within PML regions is proportional to the value of ζ. Therefore, the effectiveness of the PMLs will be increased if a larger stable value of ζ is used. From (3.77), it may be inferred that the maximum stable value of ζ will have a dependence upon the discretisation parameters, δt, δx_1 and δx_3, as well as the material system being examined since $v_{\max} = \sqrt{C/\rho}$ where C is the elastic constant in the direction of maximum velocity. Surprisingly however, the maximum stable value of ζ also has a dependence upon the thickness of the PML as well as how ζ varies through the PML. Although the derivation of universal stability limits of ζ is challenging, insight may be gained into the effect of PML thickness and functional forms of ζ by examining the stability limits numerically for a given system.

In order to systematically examine the maximum stable value of ζ we restrict its functional form to

$$\zeta_i(x_i) = \zeta_{\max}\left(\frac{|x_i - x_{i,\mathrm{PML}}|}{\Delta_{\mathrm{PML}}}\right)^a \tag{3.78}$$

where $\zeta_{\max}$ is the maximum value of ζ inside the PML, x_i is the position in the x_1 and x_3 directions, $x_{i,\mathrm{PML}}$ is the position of the boundary between the ROI and the PML and Δ_{PML} is the thickness of the PML. The shape of the ζ function may then be controlled using the parameter a, such that $a = 1$ gives a linear increase from zero at the inside edge of the PML up to $\zeta_{\max}$ at the outside edge, $a = 2$ a quadratic increase and so on. The case for $a = 0$ has been omitted since, once discretised, the PMLs cease to be perfectly matched and therefore a sudden step in ζ produces sizeable reflections from the interface between the ROI and PML which were found to be around two orders of magnitude larger than the reflected waves with steadily increasing ζ values. The maximum stable value of $\zeta_{\max}$ (with a given shape and

PML thickness) was then found by using a bisection search where the simulation was deemed to be unstable if after simulating 25 ns, by which time the initial excitation would have been absorbed by the PML, the oscillations within the simulation domain are larger than the initial excited acoustic pulse. Because instabilities within the simulation domain grow exponentially this method finds the stability limit for $\zeta_{\max}$ in the chosen system reliably. Figure 3.11 shows the maximum stable values of $\zeta_{\max}$ found for a range of values of Δ_{PML} and a for the bismuth germanate material system examined in the preceding section.

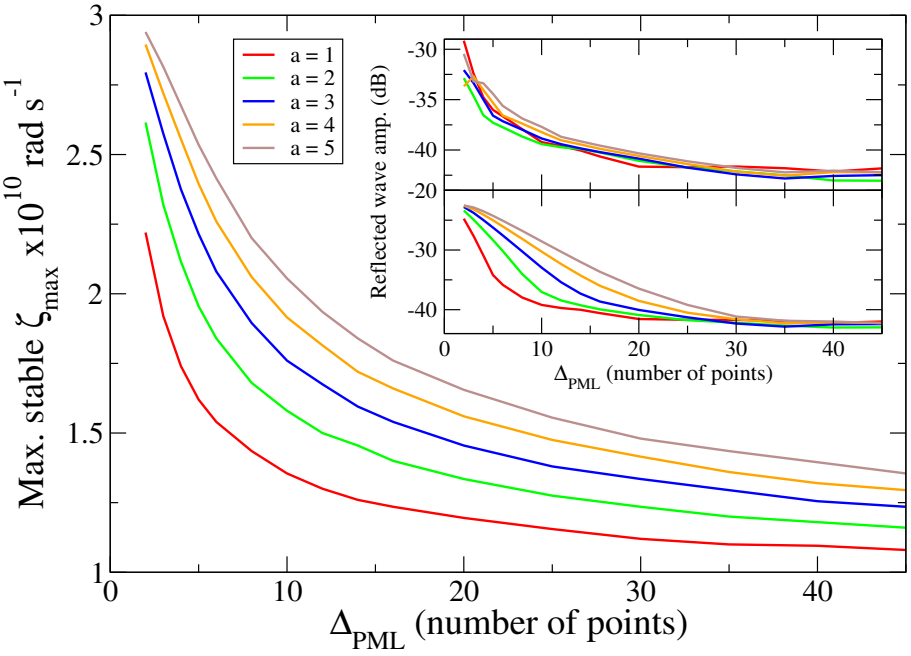

Figure 3.11: Variation in the maximum stable value of $\zeta_{\max}$ with Δ_{PML} from 2 to 45 points ($\lambda = 20$ points) and with the ζ shape factor a (as defined in equation (3.78)). *(Top inset)* The maximum amplitude of reflection from PML with varying a and Δ_{PML} using the maximum stable value of $\zeta_{\max}$ for each case, normalised to the maximum amplitude of the excited wave. *(Bottom inset)* The maximum amplitude of reflection from PML with varying a and Δ_{PML} using a constant value of $\zeta_{\max} = 1 \times 10^{10}\,\mathrm{rad\,s^{-1}}$.

To compare the effectiveness of the PMLs with different values of a, the amplitude of the wave reflected back from the boundaries was measured. This was done by performing a Fourier decomposition on the u_3 component of displacement at a point next to one of the PML boundaries over time in order to separate the reflected signal at the excitation frequency of 1 GHz from the higher frequency transients, which exist behind the excited pulse and come from the excitation being switched off suddenly after two periods. The point used was central on the x_3-axis and 40 points (6 μm) from the inside edge of one of the PML boundaries perpendicular to the x_1-axis to allow the two periods of the excited wave to pass

through the point before the reflected wave arrives back at the same point. The top inset in Fig. 3.11 shows maximum amplitudes of the reflected waves normalised to the maximum amplitude of the excited wave for different values of a and Δ_{PML}. The maximum stable value of $\zeta_{\max}$ was used in each case. For PMLs below 10-points thick, ζ with $a = 2$ is the most effective although for higher numbers of points, particularly above 20 points, all PMLs give similar performance.

Although the numerical analysis above provides a method of optimizing PML parameters for a given system, this may not be feasible for larger simulation domains, in which much longer run times would be required. Ideally, the PML parameters could be found *ab initio* for any system by applying a universal stability criterion. However, the derivation of such a criterion is challenging, and as an interim measure a more conservative choice of $\zeta_{\max}$ (i.e., much lower than those found for the example above) is likely to yield stable simulations in a wider range of simulation domains. As such, we have also examined the stability of the system considered above (using a range of shape parameters, a) when $\zeta_{\max}$ has been restricted to the much lower constant value of $1 \times 10^{10}\,\mathrm{rad\,s^{-1}}$. This value of $\zeta_{\max}$ is equal to $\frac{1}{4\delta t}$ and since $\zeta_{\max}$ is roughly inversely proportional to δt, this represents a sensible choice for a stable value of $\zeta_{\max}$ that is dependant upon the simulation parameters and the material system (as δt is material dependent). The bottom inset of Fig. 3.11 shows the amplitude of the reflected wave for different values of a and Δ_{PML}. Here we clearly see that $a = 1$ has the best performance for PMLs thinner than around 20 points. This is because higher values of a give rise to smaller ζ close to the ROI, and therefore the net attenuation of the wave within the PML is reduced.

Chapter 4

Surface acoustic wave devices

Devices which utilise surface acoustic waves (SAWs), including potential SAW-modulated quantum cascade lasers (QCLs), all have several properties in common; that is they all use an interdigitated transducer (IDT) to generate the acoustic wave, and the interactions of interest all occur near the surface of the device substrate [12]. Therefore, in order to model acoustic wave propagation through a SAW device, these points must be examined in detail. This is the focus of this chapter.

Since it is useful to have a reference as to the form of a SAW, the analytical solution to the acoustic wave equations of motion for a surface bound propagation mode [21] is presented first. This will aid in validating that any model of a SAW device does setup and maintain acoustic wave propagation of the expected form. The effect of a free surface on the acoustic wave equations of motion is examined next to determine how the FDTD bulk-acoustic-wave model may be altered to accommodate simulating surface-bound modes. This alteration comes in the form of a boundary condition of the acoustic-wave simulation domain which behaves as if it were a free surface. Finally, the structure of an IDT is considered in order to simulate the method they utilise for acoustic wave generation within the FDTD acoustic-wave model. The form of the simulated acoustic waves generated by this IDT model are compared with the analytical SAW solution to ensure that a SAW is generated, before the simulated response of a SAW device is compared with experimental results to validate the model.

4.1 Analytical solution of surface-acoustic-wave propagation

The propagation mode of a SAW is a solution of the acoustic-wave equations of motion and therefore the form of the analytical solution is reasonably easy to deduce. Finding an analytical solution of SAW propagation is advantageous as it gives a good idea of what the

form of a SAW looks like, thus can be used to check when simulating SAW devices that the obtained acoustic wave propagation is, in fact, a SAW.

Defining a free surface at $x_3 = 0$ such that the positive x_3 axes points downward into the material perpendicular to the surface, with the x_1 and x_2 parallel to the surface, the form of analytical solution of the SAW mode is

$$u_i = U_i e^{-kq_i x_3 + \mathrm{i}k(lx_1 + mx_2 - ct)} \qquad \text{for } i = 1, 2, 3, \tag{4.1}$$

where U_i is a constant defining the magnitude of the displacement in each direction, k is the wavevector, q_i is a dimensionless material dependent constant defining how *tightly* the SAW is bound to the surface, the propagation direction of the wave is at an angle θ from the x_1-axis such that $l = \cos\theta$ and $m = \sin\theta$, and c is the velocity of the propagating wave. The defining feature of the form of this wave is the exponential decay into the depth of the material which gives the surface bound propagation mode.

While (4.1) does represent the form of a surface bound propagation mode, it is incredibly general and is not an exact solution to the acoustic wave equations of motion. Clearly, an exact solution for any type of propagation mode will be intrinsically dependent upon the symmetry within the medium and can therefore not be derived without first defining the propagating material. Furthermore, any analysis to derive an exact solution will be labour intensive, as it requires the analytical solution of the equations of motion which is beyond the scope of this thesis. Details of such an analysis can be found in reference [21].

4.1.1 Solution for cubic crystals

From [21], it may be shown that within the cubic crystal class SAWs will have the form

$$\begin{aligned} u_1 &= A(e^{-qkx_3 - \mathrm{i}\phi} + e^{-q^*kx_3 + \mathrm{i}\phi})e^{\mathrm{i}k(x_1 - v_s t)}, \\ \mathrm{i}u_3 &= A(\gamma e^{-qkx_3 - \mathrm{i}\phi} + \gamma^* e^{-q^*kx_3 + \mathrm{i}\phi})e^{\mathrm{i}k(x_1 - v_s t)}, \end{aligned} \tag{4.2}$$

where A is the amplitude of the wave, $k = \frac{2\pi f}{v_s}$ is the wavevector with frequency f, q,ϕ and γ are constants depending on the material constants and v_s is the velocity of the SAW (also dependent on the material constants). Expanding the exponentials and taking only the real part, as the imaginary part is not part of the physical solution [18], gives

$$\begin{aligned} u_1 &= Ae^{-\Re(q)kz} 2\cos(\Im(q)kz + \phi)\cos(k(x - v_s t)), \\ \mathrm{i}\, u_3 &= A\Re(\gamma)e^{-\Re(q)kz} 2\cos(\Im(q)kz + \phi)\cos(k(x - v_s t)), \end{aligned} \tag{4.3}$$

where the u_3 component of displacement has a $\frac{\pi}{2}$ phase lag.

The exact form of each of the constants, as well as the velocity, will depend upon the propagation direction of the SAW, since different parts of the piezoelectric and stiffness constants will take effect. As the cubic crystal gallium arsenide (GaAs) is commonly used in SAW generation, and because of the particular relevance of GaAs to this project, this material system will be used as a test case. As already mentioned in 3.1.5, SAWs are normally launched in the [1,1,0] direction experimentally, and therefore the [1,1,0] propagation direction is used here in defining the constants for analytical SAW propagation.

From [18], the velocity of the SAW, v_s may be found by solving

$$(c_{44} - \rho v_s^2)(c_{11}c_{11}^{'} - c_{12}^2 - c_{11}\rho v_s^2)^2 = c_{11}c_{44}(\rho v_s^2)^2(c_{11}^{'} - \rho v_s^2), \tag{4.4}$$

where c_{ij} are the elements of the elastic tensor, and $c_{11}^{'}$ is given by

$$c_{11}^{'} = c_{44} + \frac{c_{11} + c_{12}}{2}.$$

This may be easily done via a numerical search over v_s since the limits on v_s are known. q is found by solving

$$(c_{11}^{'} - \rho v_s^2 - c_{44}q^2)(c_{44} - \rho v_s^2 - c_{11}q^2) + q^2(c_{12} + c_{14})^2 = 0, \tag{4.5}$$

where $c_{11}^{'}$ is defined as above. Since (4.5) is *bi-quadratic*, it may be solved easily using the quadratic formula to find q^2 and then taking the square root to find q. γ is found via

$$\gamma = \frac{(c_{12} + c_{14})q}{c_{44} - c_{11}q^2 - \rho v_s^2}, \tag{4.6}$$

and ϕ is found by

$$e^{-2\mathrm{i}\phi} = -\frac{\gamma^* - q^*}{\gamma - q}. \tag{4.7}$$

Noting that the form of these constants takes into account the propagation direction of the SAW, and therefore the piezoelectric and stiffness tensors need not be rotated from the [1,0,0] direction, the constants above come out as

$$\begin{aligned} v_s &= 2865.1\,\mathrm{ms}^{-1}, \\ q &= 0.502 + 0.477\,\mathrm{i}, \\ \gamma &= -0.681 + 1.152\,\mathrm{i}, \\ \phi &= 0.518. \end{aligned}$$

Inserting these constants into (4.3), and setting the frequency to 191 MHz to give a SAW wavelength of $\sim$15 μm at the predicted velocity of 2865 ms^{-1}, gives wave pattern shown in 4.1. For clarity, the profile of the displacement into the depth of the material has been plotted

at the maximum amplitude of displacement for the u_1 and u_3 displacement in 4.2 and 4.3 respectively. These clearly show that for a surface bound propagating wave, almost all of the waves energy is present within the top wavelength of the material.

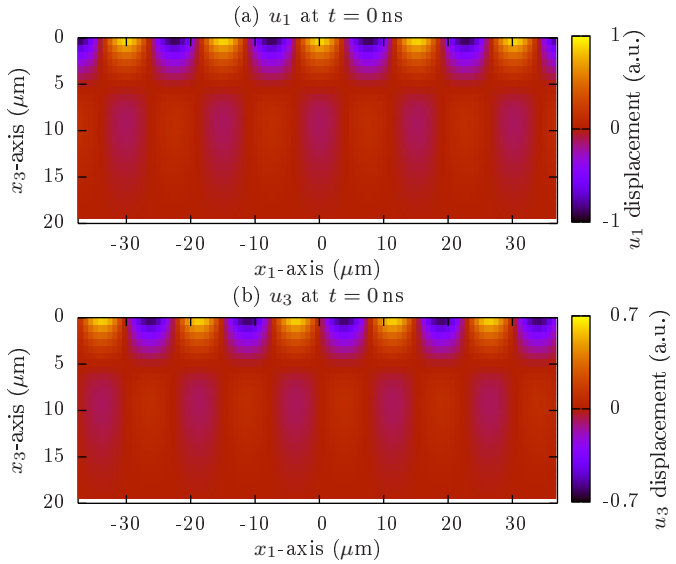

Figure 4.1: Colour plot of the u_1 and u_3 displacement for the analytical solution to the acoustic wave equations of motion for a surface bound propagation mode in GaAs (with the SAW propagation direction aligned along the [1,1,0] crystal axis).

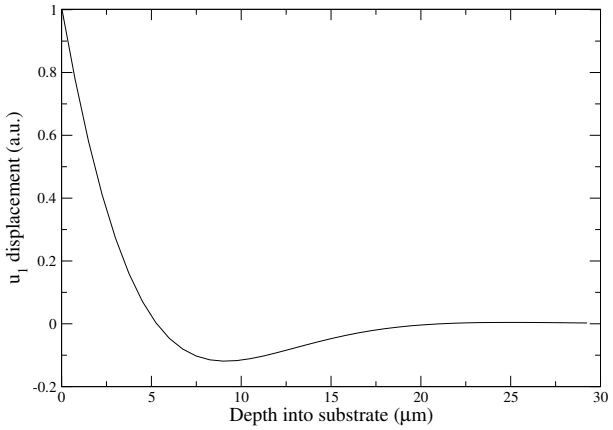

Figure 4.2: The u_1 displacement profile into the depth of the material (i.e. along x_3 axis) of the analytical solution for surface bound acoustic wave propagation (in GaAs along the [1,1,0] direction). Taken at the point of maximum u_1 surface displacement.

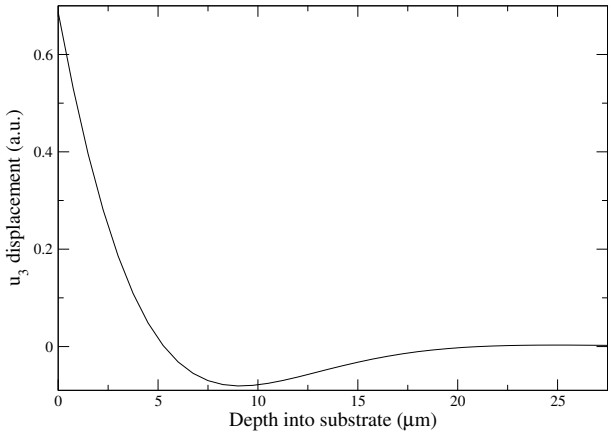

Figure 4.3: The u_3 displacement profile into the depth of the material (i.e. along x_3 axis) of the analytical solution for surface bound acoustic wave propagation (in GaAs along the [1,1,0] direction). Taken at the point of maximum u_3 surface displacement.

4.2 Modelling surface-bound acoustic-wave propagation

In order to simulate the propagation of acoustic waves along a surface of a structure, an additional boundary condition must imposed upon the simulation domain to mimic the behaviour of acoustic waves as they interact with the surface. This surface boundary condition will involve altering the equations of motion such that the propagation of acoustic energy differs at the surface from within the bulk material. The form of the surface boundary condition is therefore integral in the formation of the surface-bound propagation mode with the simulation domain. Within this thesis the widely used surface boundary condition from [21] is used, however within this section a discussion of the problems surrounding this boundary condition is also given, along with an exploration of several alternatives all of which introduce problems of their own.

4.2.1 Surface boundary condition

Before the surface boundary condition may be derived, the change this boundary makes to the simulation domain must be defined. As defined in section 4.1, the axes are orientated such that the x_3 axis points downward, into the depth of the material, perpendicular to the surface, and the x_1-x_2 plane runs parallel to the surface, with the x_1 axis being the acoustic-wave propagation direction. Imposing a flat surface boundary condition then produces the simulation domain shown in figure 4.4 with a surface on the top boundary, and the other three boundaries effectively acting as if they where an infinite bulk crystal due to the PML boundary condition.

Next the forces that act upon the surface are considered. Any force acting parallel to the surface will be unaffected by the presence of the surface and therefore the components of force, and equally stress, that act along the x_1 and x_2 direction will be the same. In the x_3 direction, the assumption is made that none of the acoustic wave energy couples from the surface into air above it; a reasonable assumption since the air is not dense, so any movement induced in the air will not have much momentum. Therefore, the differential of force, and equally the differential of the stress across the surface must be zero, i.e. with the surface situated at $x_3 = 0$,

$$\left.\frac{\partial \sigma_{i3}}{\partial x_3}\right|_{x_3=0} = 0 \qquad \text{for } i = 1, 2, 3. \tag{4.8}$$

This is the surface boundary condition, as defined in [21], which will alter the equations of

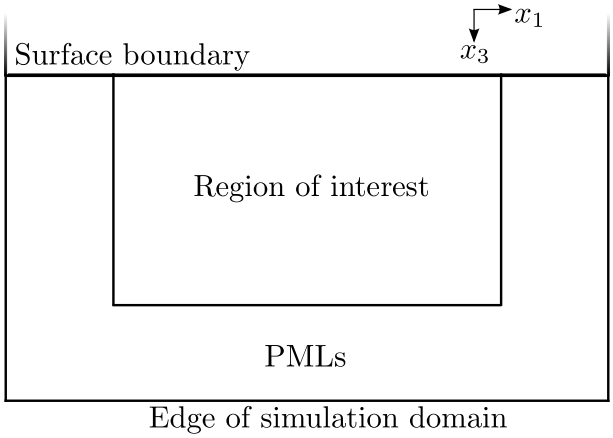

Figure 4.4: Form of the simulation domain once the surface boundary has been introduced. Three of the domain edges remain the same but the top PML has been removed and replaced with the surface boundary condition.

motion along the surface such that they become

$$\rho \frac{\partial^2 u_i(x_3 = 0)}{\partial t^2} = \frac{\partial \sigma_{ij}(x_3 = 0)}{\partial x_j} \qquad \text{for } i = 1, 2, 3; \quad j = 1, 2. \tag{4.9}$$

Within the FDTD framework, which uses the auxiliary variable v instead of σ, the assumption is made that $\frac{\partial \mathsf{v}}{\partial x_3}$ is also zero at the surface. This is reasonable since σ is the time differential of v, and there is no reason why the differentiation will not commute. The implementation of the surface boundary within the FDTD framework is therefore trivial since the calculation of v is unchanged as there are no u-points with are *missing* from the simulation domain, only those which lie along the surface, and the calculation of u along the surface simply involves setting the $\frac{\partial \mathsf{v}}{\partial x_3}$ equal to zero.

4.2.2 Adapting PML boundaries to include the surface boundary condition

The surface boundary condition alters the acoustic wave equations of motion at the surface of the strucutre being simulated. Since this surface extends to the edge of the simulation domain, and therefore into the PML region as in figure 4.4, the form of the equations within the PML must be altered to include the surface boundary condition. This involves performing an analytic continuation of spatial variables onto the complex, as in 3.3.1, on the equation of motion at the surface boundary.

Applying the surface buondary condityion to the equations of motion gives

$$\rho \frac{\partial u_i}{\partial t} = \frac{\partial v_{i1}}{\partial x_1} \qquad \text{for } i = 1, 2, 3, \tag{4.10}$$

where $i = 1, 2, 3$ for the rest of the derivation and so shall no longer be stated. As stated within section 4.2.1, the auxiliary variable v only exists below the surface and is not altered by the surface boundary condition. Transforming to the time domain, such that $\mathbf{u}(t) \rightarrow \mathbf{U}(\omega)$ and $\mathbf{v}(t) \rightarrow \mathbf{V}(\omega)$,

$$\rho \mathrm{i} U_i = \frac{\partial V_{i1}}{\partial x_1}, \tag{4.11}$$

and performing the substitution of variables, $x_i \rightarrow (1 + \mathrm{i}\frac{\zeta_i}{\omega})x_i$, gives

$$-\mathrm{i}\omega\rho U_i = \frac{1}{1 + \mathrm{i}\frac{\zeta_1}{\omega}} \frac{\partial V_{i1}}{\partial x_1}. \tag{4.12}$$

Multiplying by $(1 + \mathrm{i}\frac{\zeta_1}{\omega})$,

$$-\mathrm{i}\omega\rho U_i + \rho\zeta_1 U_i = \frac{\partial V_{i1}}{\partial x_1}, \tag{4.13}$$

and transforming back to the time domain gives

$$\rho \frac{\partial u_i}{\partial t} = \frac{\partial v_{i1}}{\partial x_1} - \rho\zeta_1 u_i. \tag{4.14}$$

Since this formalism does not contain any addition auxiliary variables, the implementation of the surface boundary inside the PML region is relatively simple; that is, it involves only adding one extra term to the equations of motion at the surface. However, as discussed in 3.3.3, (4.14) infers a dependence on $\mathbf{u}$ at a time step where it is not calculated, i.e. $\mathbf{u}(t+\frac{1}{2})$. This may be dealt with by utilising the same assumption as in 3.3.3, that the discretisation grid should be tight enough that, for the implementation of the PMLs, adjacent time steps may be assumed to be approximately equal, i.e. $\mathbf{u}(t + \frac{1}{2}) \approx \mathbf{u}(t)$.

4.3 Interdigitated transducers

4.3.1 Real-world SAW excitation

SAWs are excited using an IDT, which consist of a series of metal finger electrodes which are deposited onto the surface of a piezoelectric substrate, with different fingers connected to different voltage supply lines. The most common configuration for an IDT is to have alternate fingers connected to alternate supply lines as shown in 4.5, although some devices have proceeding fingers connected to the same side and others have more than two supply lines.

Whilst the complexity of IDTs varies, they all reply on the piezoelectric effect to generate SAWs. Consider two electrodes connected to constant, positive and negative voltages, which are bought into contact with the surface of a piezoelectric material. The electric field around

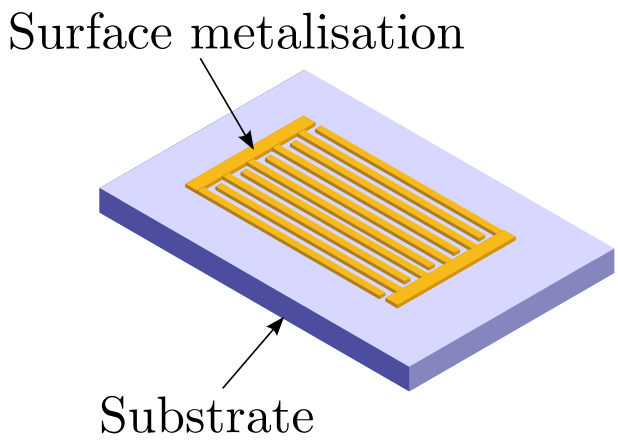

Figure 4.5: Diagram of the form of a basic IDT made by depositing metal onto the surface of a piezoelectric substrate.

the two electrodes which propagates through the material will induce localised strain within the material via the piezoelectric effect. This localised strain, which will initially be directly underneath the electrodes, will propagate through the material at the speed of the sound as the material adjusts to the locally induced strain, and reaches some equilibrium state. If the electrodes are now connected to either side of a sinusoidally varying voltage source, then the induced strain will now oscillate, from tensile to compressive, and will generate a pressure wave the will propagate radially outwards through the material at the speed of sound.

Using multiple pairs of fingers, with each alternate finger connected to a different side of the alternating voltage source, then multiple pressure waves will be launched radially from underneath each finger pair. If each finger is regularly spaced by a half-wavelength of the propagating pressure waves within the material then there will be a regular interference pattern between waves as shown in 4.6. Destructive interference below the fingers causes a small acoustic wave amplitude to be found below the surface. Along the surface the propagating waves constructively interfere such that as a wave passes under the next finger-pair its amplitude is added to the strain induced by the finger-pair, effectively amplifying the acoustic wave amplitude as it passes underneath the IDT structure. This causes a large wave amplitude along the surface. It also means that the amplitude of the wave emitted by the IDT does not reach a steady state until the propagating wave generated at one end of the IDT has passed through the entire length of the structure — something which is shown by the FDTD model in 4.3.3.

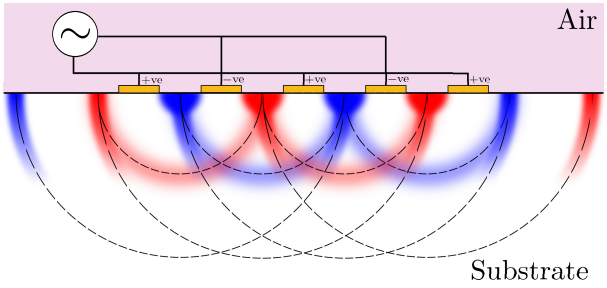

Figure 4.6: Cross-section of the interference pattern of the acoustic waves below the substrate surface, underneath an IDT. The red and blue areas indicate regions of induced stress of different signs which propagate radially downwards from in between each finger pair. (The peak amplitudes of the radially propagating waves are shown by dashed lines for clarity.) With the periodicity of the fingers at $\frac{\lambda}{2}$ of the applied oscillating frequency (as shown here), each peak will constructively interfere with the peak generated underneath the adjacent IDT pair close to the surface. This produces an almost entirely surface bound acoustic wave which propagates from each end of the IDT.

4.3.2 Modelling IDTs

Before a model of IDTs can be put together, the assumptions that are to be used must be defined. Since the FDTD acoustic wave model defined previously already assumes that the waves are invariant parallel to the wave front, the assumption will also be used here. In fact, this assumption comes from the fact the majority of IDT structures are invariant along the length of the fingers and therefore produce waves with approximately flat wavefronts. While this excludes the possibility of simulating more complicated IDT types, these are not of interest within the scope of this thesis so can be excluded.

Secondly, the electronic properties of the IDT are considered. Because the excitation frequencies for the IDTs are often in radio-frequency range (or even higher, up to several GHz), in order to determine the exact form of the electric field around the IDT structure and transmission-line style analysis would be required. This sort of analysis would not only significantly increase the complexity of SAW devices when included within the FDTD simulation and would mean that the electric field around each finger would not need to be constant down the length of each finger, therefore meaning that the excited acoustic waves would not be invariant along the wave front. As the assumption of invariance along the acoustic wave front has already been made, and to reduce the complexity of the simulation, it is also assumed that the electric field around the IDT is invariant along the length of each finger.

This assumption is equivalent to assuming that the movement of charge within the IDT is adiabatic compared to the excitation frequency,[1] which infers that the voltage on the IDT fingers changes instantaneously with the applied voltage. Viewing the IDT as capacitor, this assumption can be taken further to say that the voltage across the width of each finger will also be constant, which reduces determining the electric field around the IDT structure to solving Poisson's equation around some fixed voltages.

Finally, the acoustic properties of the IDT and considered. Since the depth of the metallisation that makes up the IDT structure is several orders of magnitude smaller than the wavelength of the SAW begin generated for *lower* frequencies, it can be assumed that metal will have no mechanical effect on the propagating wave. When the wavelength of the SAW is short enough that it becomes comparable to the metallisation depth the problem of determining how the acoustic wave enters the IDT fingers becomes complicated since the metal will either be amorphous of polycrystalline (depending on the conditions when depositing). However, since the lowest achievable SAW wavelengths are around 640 nm [13] with IDT metallisation depths of the order of 40 nm, the effect of the acoustic wave entering the metallised region can be ignored. What cannot be ignored at shorter wavelengths however, is that the mass of the metal on the surface of the substrate becomes a significant factor in how the acoustic wave passes through the material. This *mass loading* effect is discussed in 4.3.6.

Now that all assumptions have been defined the IDT model can be constructed. Since all mechanical effects are to be ignored, the model of the IDT consists of artificially introducing a potential within the simulation domain at every point in time. As Poisson's equation is already solved at every time step within the FDTD model already, implementing the IDT model simply involves fixing the potential within Poisson's equation.

Fixing the potential in Poisson's equation

As derived in 3.2.3, the discretised Poisson's equiation in two dimensions takes the form

$$\mathbf{A}\phi = \rho \tag{4.15}$$

where the matrix $\mathbf{A}$ is banded and must be diagonalised (or reduced to a triangular matrix and back substitution used) in order to find the potential, ϕ, resulting from some charge distribution, ρ. For the potential to be fixed, i.e. the potential at some point to remain the same before and after diagonalisation, the relationship between ϕ and ρ must remain the same at that point after diagonalisation. Since the relationship between ϕ and ρ changes

[1] This will clearly hold true more for lower frequencies than high frequencies.

during diagonalisation to minimise the off diagonal elements of $\mathbf{A}$ to zero, the only way to ensure that their relationship remain constant is to make the rows of $\mathbf{A}$ which correspond to the points of fixed potential diagonal to begin with. This is done by setting the off diagonal elements to zero within the row which corresponds to the point in space where the potential is to be fixed, i.e. taking the original form of the two-dimensional Poisson's equation from 3.2.3,

$$\begin{pmatrix} a_{1,1} & b_{2,1} & c_{1,1} & & & & & & \\ b_{2,1} & a_{2,1} & \ddots & \ddots & & & & & \\ c_{1,1} & \ddots & \ddots & \ddots & \ddots & & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & & \\ & & c_{i-ni,j} & b_{i-1,j} & a_{i,j} & b_{i,j} & c_{i,j} & & \\ & & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & & \ddots & \ddots & \ddots & b_{ni,nj} & \\ & & & & & c_{ni-ni,nj} & b_{ni,nj} & a_{ni,nj} & \end{pmatrix} \begin{pmatrix} \phi_{1,1} \\ \phi_{2,1} \\ \vdots \\ \vdots \\ \phi_{i,j} \\ \vdots \\ \vdots \\ \phi_{ni,nj} \end{pmatrix} = \begin{pmatrix} \rho_{1,1} \\ \rho_{2,1} \\ \vdots \\ \vdots \\ \rho_{i,j} \\ \vdots \\ \vdots \\ \rho_{ni,nj} \end{pmatrix}, \quad (4.16)$$

where a, b and c are the five non-zero diagonals (the zero elements are not shown), if the potential is to fixed at the point i, j then (4.16) becomes

$$\begin{pmatrix} a_{1,1} & b_{2,1} & c_{1,1} & & & & & & \\ b_{2,1} & a_{2,1} & \ddots & \ddots & & & & & \\ c_{1,1} & \ddots & \ddots & \ddots & \ddots & & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & & \\ & & 0 & 0 & a_{i,j} & 0 & 0 & & \\ & & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & & \ddots & \ddots & \ddots & b_{ni,nj} & \\ & & & & & c_{ni-ni,nj} & b_{ni,nj} & a_{ni,nj} & \end{pmatrix} \begin{pmatrix} \phi_{1,1} \\ \phi_{2,1} \\ \vdots \\ \vdots \\ \phi_{i,j} \\ \vdots \\ \vdots \\ \phi_{ni,nj} \end{pmatrix} = \begin{pmatrix} \rho_{1,1} \\ \rho_{2,1} \\ \vdots \\ \vdots \\ \rho_{i,j} \\ \vdots \\ \vdots \\ \rho_{ni,nj} \end{pmatrix}. \quad (4.17)$$

Since $\phi_{i,j}$ and $\rho_{i,j}$ will remain in a constant ratio the value of $a_{i,j}$ is not important, although from an implementation point of view it is simpler to give it the same form as the rest of the diagonal. The potential at the point i, j may then be fixed to some arbitrary potential, V, by setting $\rho_{i,j} = \frac{V}{a_{i,j}}$.

While making this change to the matrix $\mathbf{A}$, implies that it is no longer Poisson's equation that is being solved, importantly it is akin to solving Poisson's equation around the area of fixed potential with fixed-potential boundary conditions, therefore Poisson's equation is solved normally in the areas where the potential is not fixed. Within the areas of fixed potential Poisson's equation is clearly no longer solved, so when fixing the potential in this

manner care must be taken to ensure that charge is not *placed* within these areas as it would not result in the potential field expected.

Solution of Poisson's equation around an IDT structure

Since the potential within Poisson's equation may be fixed at as points as desired, and the solution to Poisson's equation has been implemented as part of the FDTD simulation of acoustic wave, finding the potential around an IDT structure simply involves defining the positions of the IDT fingers and potential on each finger. Furthermore, as already stated, since the thickness of the IDT metallisation is several orders of magnitude smaller than the width of each finger, the IDT structure can be assumed to be one grid point in height.

Figure 4.8 shows the potential found around a 5-finger-pair IDT sturcutre in the GaAs material system with each finger being 3.75 μm wide and separated by 3.75 μm, as shown in 4.7, and alternate fingers connected to +10 V and -10 V.

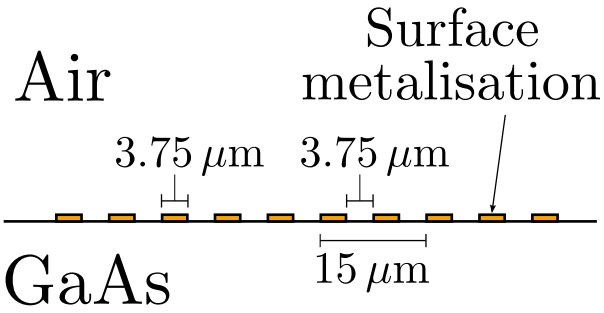

Figure 4.7: Schematic diagram of the IDT simulated in figure 4.8.

4.3.3 Excitation of FDTD model via IDTs

As described in 4.3.2, the basis for the model of an IDT involves fixing the potential within Poisson's equation to find the potential profile around the IDT structure. As Poisson's equation is solved at every time step within the framework of the FDTD simulation, using the IDT model to excite acoustic waves within this framework is trivial and involves applying the changes to the Poisson solver outlined in section 4.3.2. Since the time varying potential applied to IDT devices experimentally is sinusoidal, it is defined as

$$\phi_{\mathrm{IDT}} = A * \sin(2\pi \cdot f \cdot it \cdot \delta t) \tag{4.18}$$

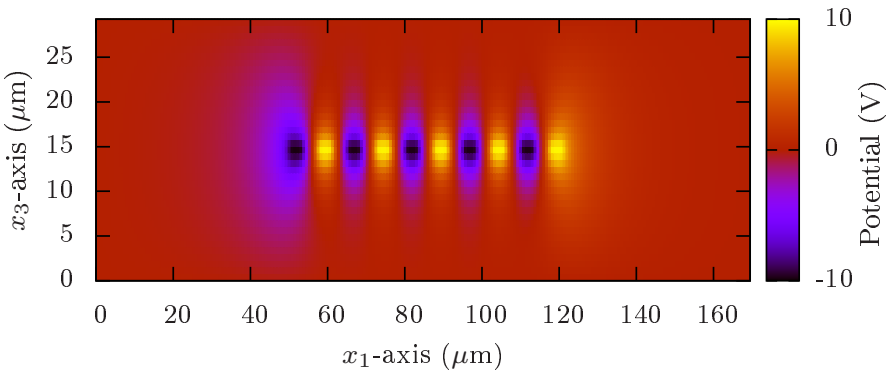

Figure 4.8: The potential profile around a 5-finger-pair IDT placed on the surface of GaAs with finger widths of 3.75 μm and spaces between fingers of 3.75 μm (giving a SAW emisson wavelength of $15m\mu$m). The boundary between the GaAs and the air is at $x_3 = 15\,\mu$m.

where ϕ_{IDT} is the potential applied to one finger, A is the amplitude of the applied potential, f is oscillation frequency of the applied potential, it is the number of the current time step and δt is the temporal discretisation step. Because the method is general however, the applied bias may be of any form desired or applied to metallised areas not in the form of IDTs, as discussed in 5.1.1 which is concerned with applying a fixed (in time)potential to metallised surfaces.

Result of IDT excitation

The oscillating potential form a 5-finger-pair IDT with a SAW wavelength of 15 μm (as shown in figure 4.7) was used to excite the FDTD acoustic-wave model using the GaAs material system, with the x_1 axis aligned along the [1,1,0] crystal axis (i.e. the wave propagation direction is along [1,1,0]). Since the speed of sound in the [1,1,0] direction in GaAs is approximately 2800 ms^{-1}, the frequency of a 15 μm wavelength SAW will be $\sim$ 187 MHz. (Experimentally, the *accepted* frequency is 188 MHz, giving a velocity of 2820 ms^{-1}, but this value varies depending on which source is quoted.) With this in mind, the frequency of the potential oscillations is set to 187 MHz as this should be the resonant frequency of the IDT. The spatial discretisation steps, δx_1 and δx_3, where both set to 0.75 μm, i.e. $\delta x_i = \frac{\lambda}{20}$, and the temporal discretisation step, δt, was set to 0.05 ns, i.e. $\delta t = \frac{\tau}{100}$ where τ is the time period of the SAW.

Figures 4.9 and 4.10 show the normalised u_1 and u_3 displacements over a range of time intervals. (The IDT fingers are between 50- and 100 μm on the x_1 axis.) Clearly an acoustic wave is launched from either end of the IDT structure, which propagates outwards along the

surface (located at $x_3 = 0\,\mu$m). Furthermore, the acoustic wavelength is $\sim 15\,\mu$m and most of the acoustic wave energy is contained within one wavelength below the surface, strongly indicating that a surface acoustic wave has been induced.

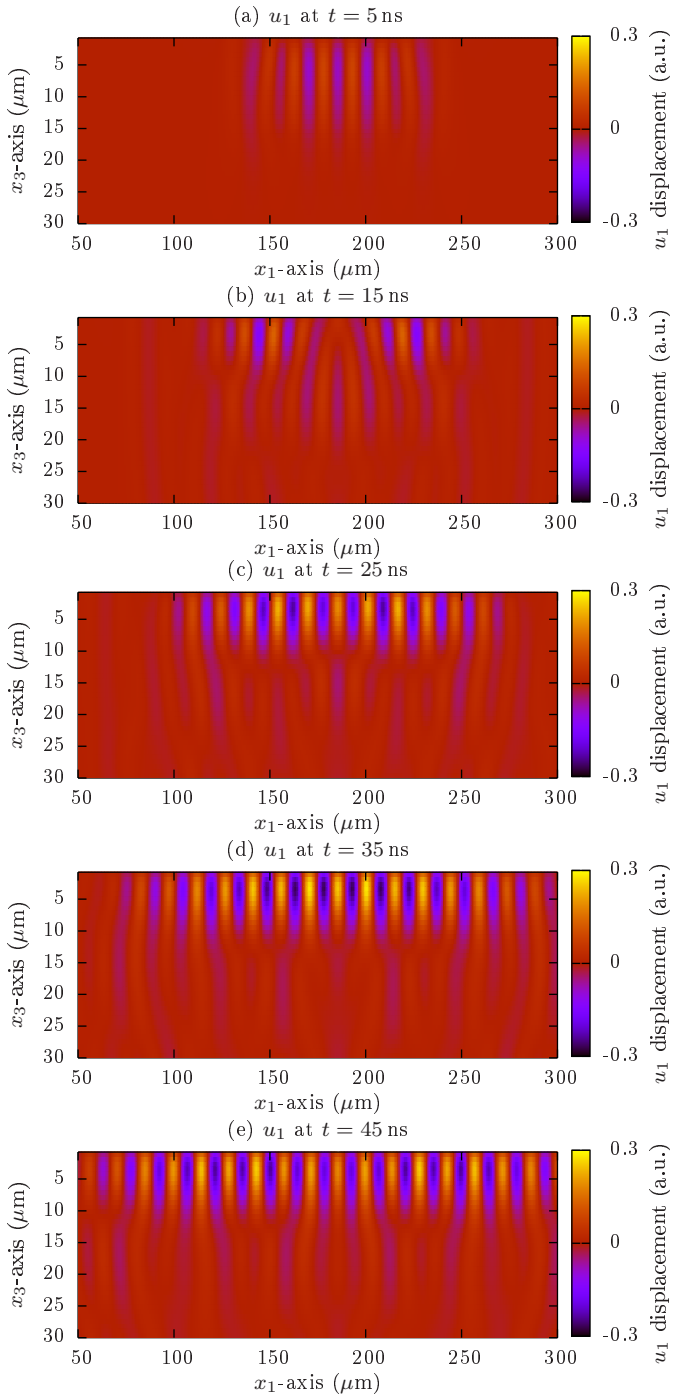

Figure 4.9: Induced u_1 displacement at various time intervals for a sinusoidally varying potential around a 5-finger-pair IDT, with finger widths and spacings of 3.75 μm, giving a SAW wavelength of 15 μm at a frequency of 188 MHz.

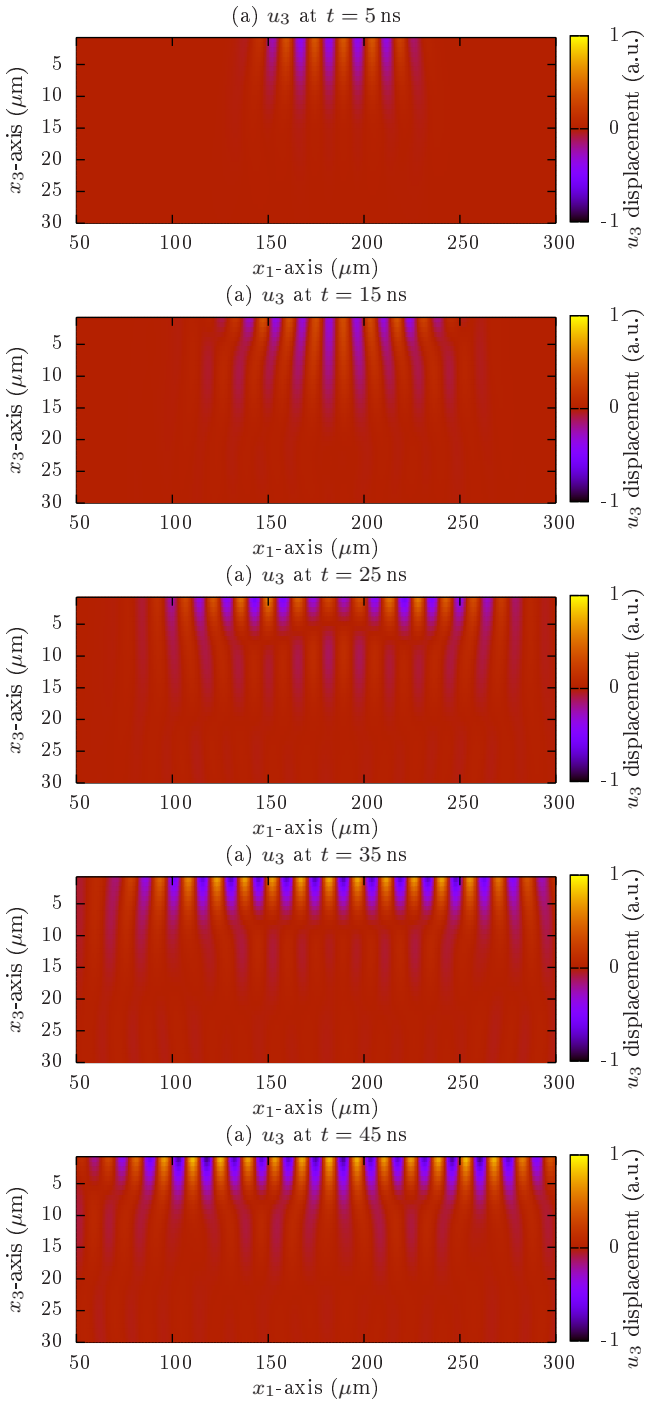

Figure 4.10: Induced u_3 displacement at various time intervals for a sinusoidally varying potential around a 5-finger-pair IDT, with finger widths and spacings of 3.75 μm, giving a SAW wavelength of 15 μm at a frequency of 188 MHz.

Comparison with analytical excitation

Comparing the analytical solution for SAW propagation with the acoustic wave excited via the model of IDTs, one can see that the two are clearly not identical. The main difference is the depth of the mode confinement; for the analytical solution the majority of the mode is confined within $\sim 5\,\mu$m ($\frac{\lambda}{3}$) below the surface, whereas the IDT excitation produces a mode which is mostly confined within the top $10\,\mu$m ($\frac{\lambda}{2}$). Furthermore, the non-zero displacements, u_1 and u_3, do not have the same shape with the IDT excitation as they do in the analytical excitation. This can be seen from the comparison of displacements profiles over depth, taken at the maximum amplitude of displacement, in figures 4.11 and 4.12 for u_1 and u_3 respectively.

While the two methods do not produce modes which are identical however, they do produce modes which are similar. Both have almost all of the wave energy within the top wavelength of the surface, and both have the u_3 phase-lagging by $\frac{\pi}{2}$ from u_1. Furthermore, real-world IDTs are not expected to produce a *pure*, surface-bound propagating mode, but a mixture of surface- and bulk- propagating modes. In fact, there are some applications which make use of this fact, using specially designed IDTs to specifically excite modes other than surface-bound, such as leaky-SAWs and shallow bulk acoustic waves, as discussed by [15].

The fact that the two methods do not produce identical modes is not a problem, but happens because real-world IDTs do not produce a pure SAW mode, therefore making the solution found via the IDT method more physically realistic.

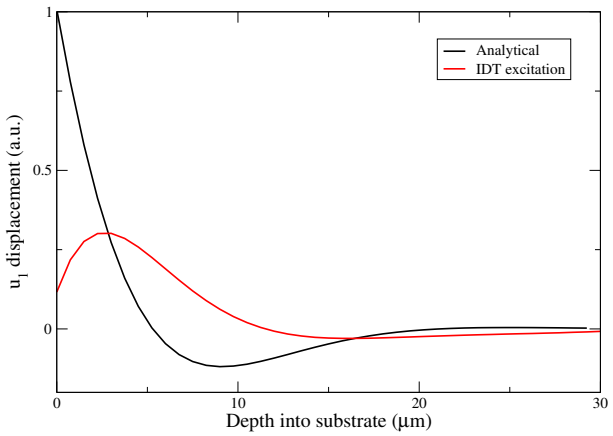

Figure 4.11: Comparison of the u_1 displacement depth profile for the analytical solution for a SAW mode and IDT excitation of FDTD model, both done in GaAs material system with a [1,1,0] propagation direction. Amplitudes are normalised to the maximum displacement.

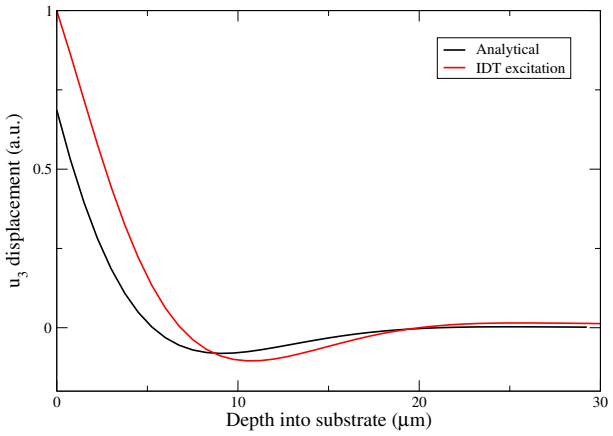

Figure 4.12: Comparison of the u_3 displacement depth profile for the analytical solution for a SAW mode and IDT excitation of FDTD model, both done in GaAs material system with a [1,1,0] propagation direction. Amplitudes are normalised to the maximum displacement.

4.3.4 Transmitting and receiving IDT pairs

Real-world SAW devices are most commonly fabricated with a pair of IDTs separated by some distance as shown in figure 4.13. This acts as a form of fail-safe, to check that the IDTs are functioning properly by testing the S-parameters of the overall device as in figure 4.14. Despite the fact that the technology to fabricate IDTs at microwave frequencies is well established, fabrication defects are still common, owing to the long, thin geometry of each finger as well as thin spacing between fingers and the shear number of fingers on each device. Furthermore, the devices themselves are very fragile, with a short circuit between to the two sides of the IDT structure enough to stop the device working all together. It is therefore very attractive to be able to test the operation of the SAW device before any more complicated devices are placed in between the two IDTs, for example, as is the case within this thesis, a QCL.

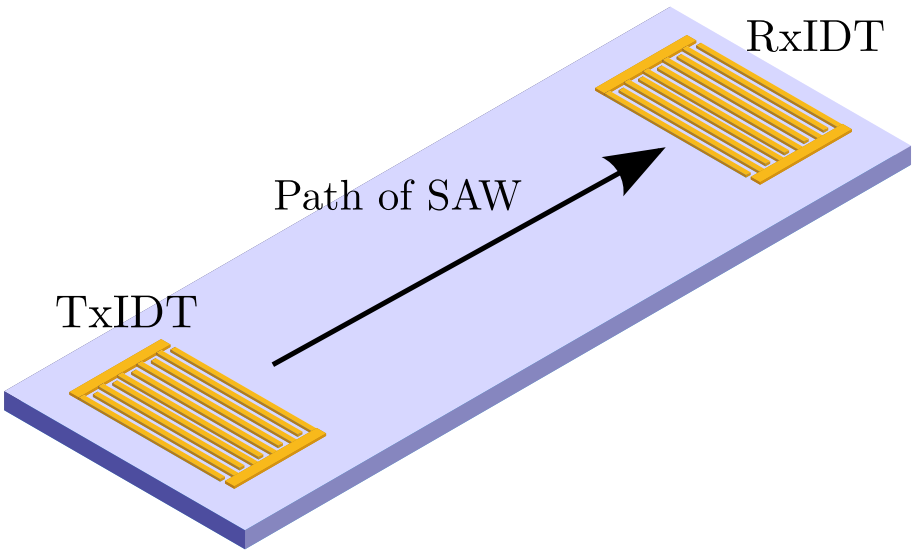

Figure 4.13: Schematic diagram of a SAW device with a transmitting IDT (TxIDT) for SAW generation and receiving IDT (RxIDT) for SAW detection.

Since a model SAW generation via IDTs has presently been developed, the best form of model validation is to compare the theoretical predictions with the experimental results for a transmitting and receiving IDT pair. Before this can be done, however, the physics of an IDT *receiving* a SAW must be considered. Furthermore, the geometry of the real-world device must be considered; that is, with the inclusion of the PML the simulated devices will effectively be sat atop an infinite substrate, where as in reality the device has well defined edges which the SAW will interact with.

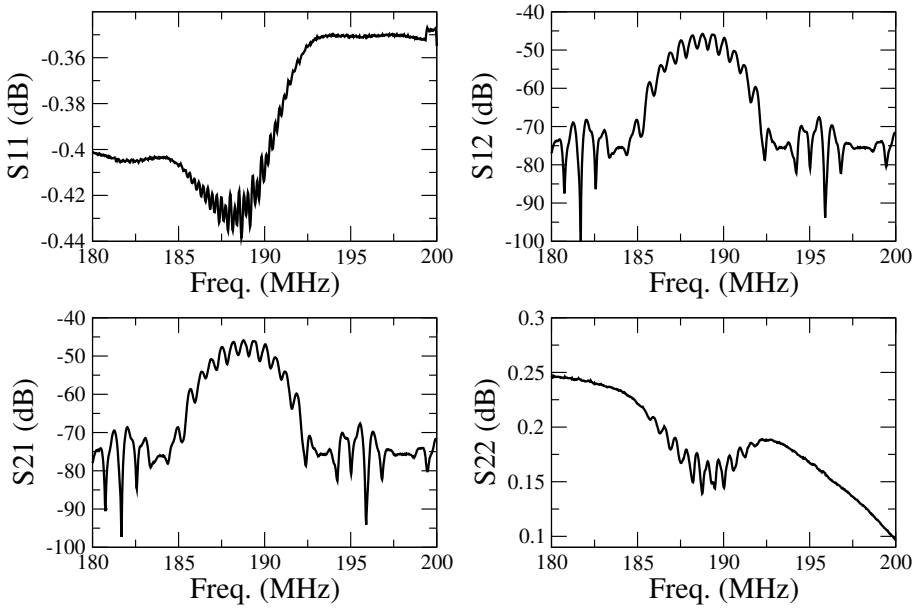

Figure 4.14: Experimentally measured S-parameters for a SAW device with 40-finger-pair IDTs with a design wavelength of 15 μm, propagating in the [1,1,0] direction on GaAs. (Device shown fabricated and measured by Dr M. Salih.)

Receiving IDTs

A travelling SAW carries with it a travelling electric field wave, generated via the piezoelectric effect. As a SAW travels underneath any metallised surface, its electric field will induce a movement of charge within the metal. If the SAW travels beneath an IDT structure, then a different charge will be induced on each finger (depending on where each finger happens to be within the SAW wavelength), and these will ad either constructively or destructively depending on the dimensions and structure of the IDT, causing an imbalance charge on each side and therefore a voltage across the IDT. If the grating length of the IDT matches half a wavelength of the SAW, i.e. if the IDT would emit at the same frequency as the SAW, then the charge on each finger will add constructively and a large voltage will be detectable across the IDT structure.

Despite this relatively simple action of a receiving IDT, modelling the exact movement of charge within the IDT metal becomes quite complicated since it involves considering the electromagnetic interaction between the SAW electric field and the IDT (i.e. transmission line theory). Such an analysis is beyond the scope of the current work, especially since modelling receiving IDT is not a necessary component of the modelling the interaction between SAWs and quantum cascade lasers. However, the response of a receiving IDT can be approximated by assuming the overall response will be proportional to the sum of the SAW voltages underneath each of the IDT fingers. This should hold true since the pattern of constructive/destructive interference is caused the geometry of the IDT, not its electrical properties. Furthermore even if a full electronic model of the IDT were developed, this would still not necessarily be an accurate description of the measured voltage across the IDT, since it would not include the effect of the bond wires connected to the IDT which would almost certainly not be impedance matched. Therefore, the response of the simulated receiving IDT, R_{IDT}, may be found by taking the mean average of the voltages at every point, on every finger of the IDT i.e.

$$R_{\mathrm{IDT}} = \frac{\sum_i^{n_{\mathrm{IDT}}} V_{\mathrm{SAW}}(i)\delta x_1}{n_{\mathrm{IDT}}}, \tag{4.19}$$

where n_{IDT} is the number of grid points within the metallised fingers of the IDT, i is an index over all the metallised grid points within the IDT and $V_{\mathrm{SAW}}(i)$ is the voltage underneath the i^{th} point of metallisation in the IDT. The dimensions of the IDT are considered by the inclusion of the δx_1 term.

Using this approximation of summing the SAW voltage under each finger, one may find the response of the receiving IDT at every point in time. By taking the maximum steady-

state amplitudes of these responses over a range of frequencies, the overall frequency response of an transmitting/receiving IDT pair may be determined. Figure 4.15 shows the frequency response for a pair of 20-finger-pair transmitting and receiving IDTs both with a SAW wavelength of 15 μm, in the GaAs [1,1,0] material system, separated by a distance of 450 μm. This

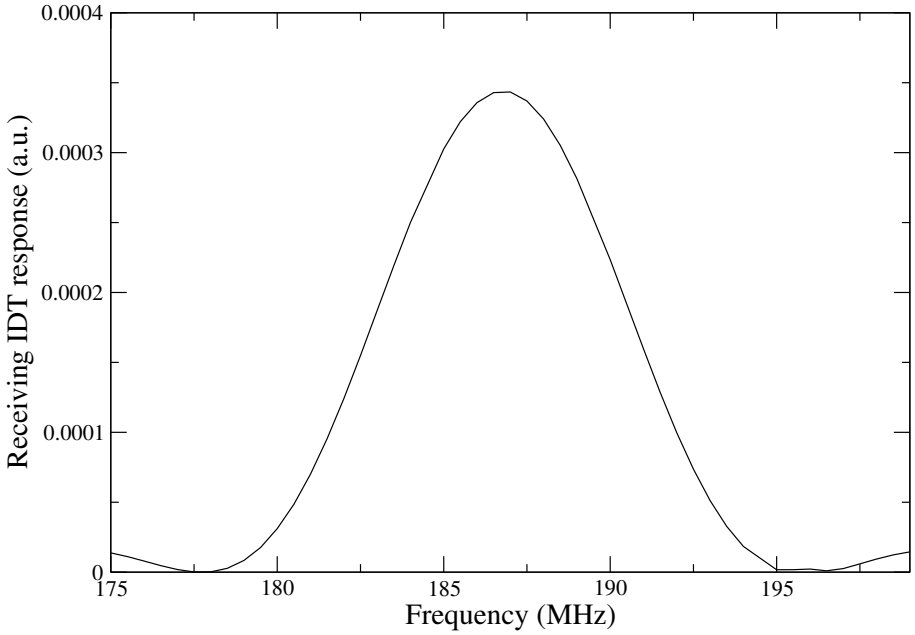

Figure 4.15: Receiving IDT response for a pair of 40-finger-pair transmitting/receiving IDTs with SAW wavelength of 15 μm, in the GaAs material system with the propagation direction being [1,1,0], and separated by a distance of 450 μm. The applied voltage on each side of the transmitting IDT was ± 10 V.

shows that the resonant frequency of the IDT pair is at 186.75 ± 0.25 MHz which, given the wavelength of SAW emission, makes the velocity of the SAW 2801.25 ms^{-1}.

As discussed in 4.3.1, the response of an IDT pair should not reach a steady state until the acoustic wave excited from the first finger of the transmitting IDT has reached the last finger of the receiving IDT. With the current SAW device, where the length of each IDT is $20 \times 15\mu\mathrm{m} = 300\mu\mathrm{m}$ and the separation between IDTs is 450 μm, this steady-state *distance* is $2 \times 300 + 450 = 1050\mu$m. Therefore, with the SAW velocity calculated from the simulated resonance of the device, the steady-state time of the device is $\frac{1050\mu\mathrm{m}}{2801.25\mathrm{ms}^{-1}} = 374.8$ ns. Figure 4.16 shows the time dependent response of the receiving IDT at the resonant frequency with the steady-state time marked on. Clearly the steady-state time calculated from the resonance of the device is correct, indicating that SAW velocity calculated from this method

is also correct. Furthermore, since the resonant frequency of the device, and therefore the calculated velocity, is very close to experimentally measured values for GaAs with SAW propagation in the [1,1,0] direction, this indicates that simulation is a good approximation to real SAW propagation. Interestingly, when the time dependent responses of resonance are

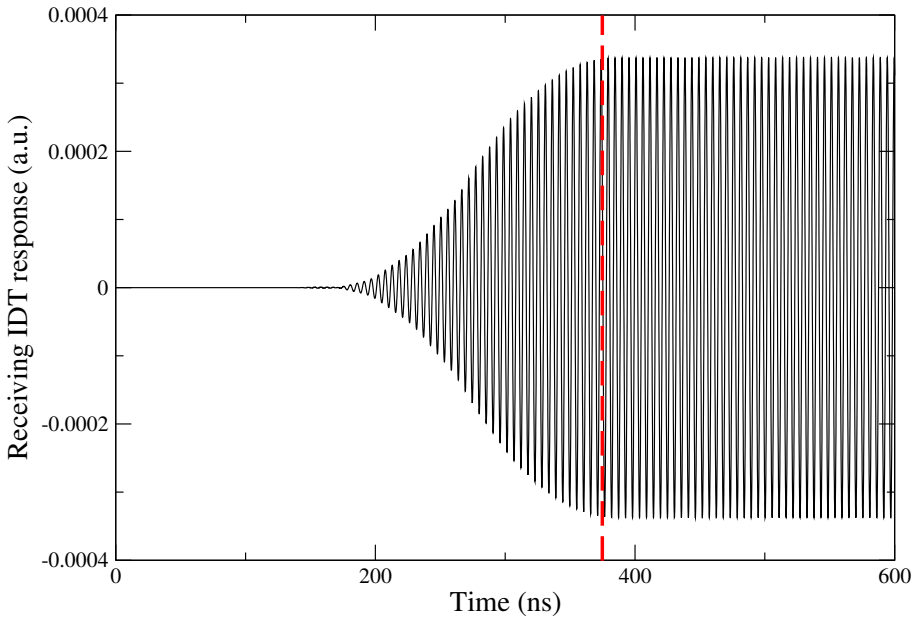

Figure 4.16: Time dependent response of the receiving IDT for the same SAW device simulated in 4.15 with the transmitting IDT driven at 187 MHz. The calculated steady-state time for the device of 374.8 ns is marked by the red, dashed line.

examined, it is found that the steady-state response is in fact lower than the initial signal. Figures 4.17 and 4.18 show the responses with the transmitting IDT driven at 181 MHz and 178 MHz respectively. Clearly the steady-state time of the response does not change, which is expected since the acoustic velocity is not frequency dependent. The steady-state value is lower than the initial pulse because just off resonance the receiving IDT fingers will be a combination of constructive and destructive interference, and it just so happens at these frequencies that the destructive interference does not overcome the constructive until a large portion of the SAW has entered the receiving IDT.

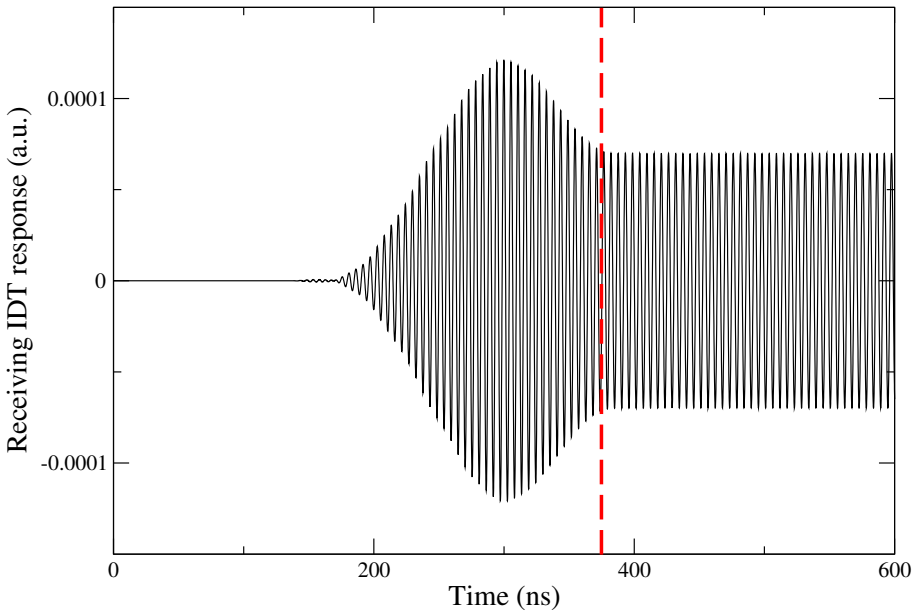

Figure 4.17: Time dependent response of the receiving IDT for the same SAW device simulated in 4.15 with the transmitting IDT driven at 181 MHz. The calculated steady-state time for the device of 374.8 ns is marked by the red, dashed line.

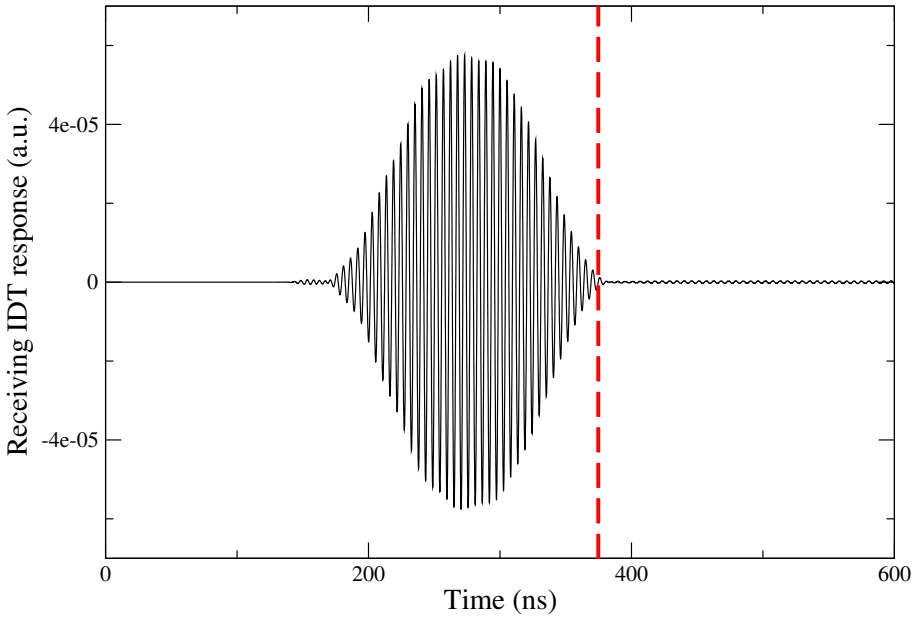

Figure 4.18: Time dependent response of the receiving IDT for the same SAW device simulated in 4.15 with the transmitting IDT driven at 178 MHz. The calculated steady-state time for the device of 374.8 ns is marked by the red, dashed line.

As discussed above, the absolute value of the simulated response from the receiving IDT has no real meaning since it is found via an approximate method. However, the overall transmission through the device may (in decibels) be determined by finding the input power using the same method for finding the response of the receiving IDT on the transmitting IDT. The ratio of the two may then be used to the find the response in decibels. Figure 4.19 shows the decibel response for the same SAW device simulated above. From this, the characteristic sinc function-like response of the SAW device which both predicted analytically [15] and measured experimentally; another good indicator that the FDTD model of SAW devices is accurate.

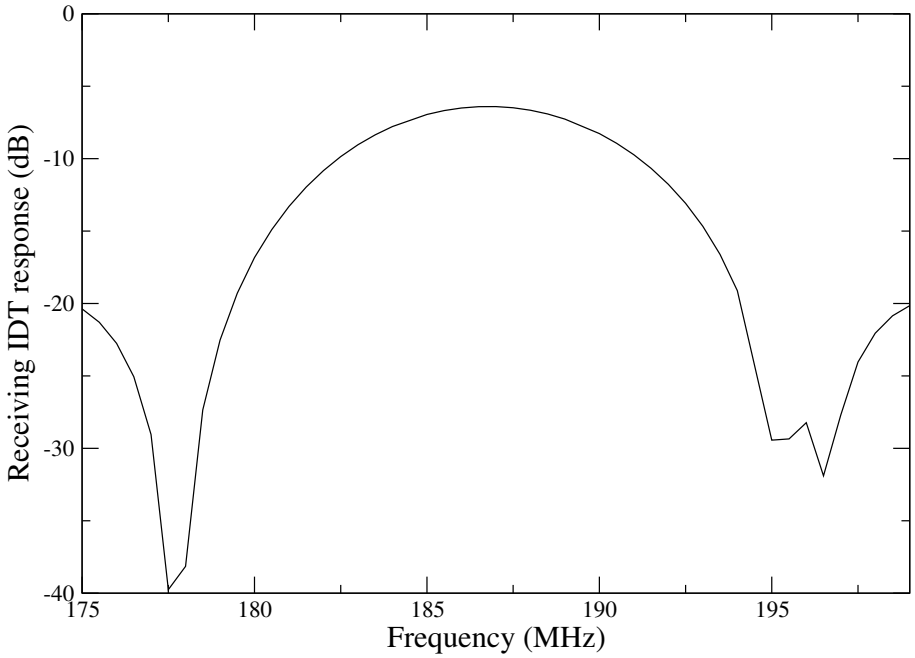

Figure 4.19: Receiving IDT response in decibels for a pair of 40-finger-pair transmitting/receiving IDTs with SAW wavelength of 15 μm, in the GaAs material system with the propagation direction being [1,1,0], and separated by a distance of 450 μm. The applied voltage on each side of the transmitting IDT was ± 10 V.

Devices edges

Within the simulation domain, three of the boundaries of the modelled SAW devices are made of PMLs. These act not only to stop artificial reflections off these boundaries, but also to remove the need for the simulation domain to be the size of the actual devices, by making these three boundaries act like an infinitely extending substrate. (This is particularly

important to reduce the depth of the simulation domain as the scalability of the Poisson solver within the simulation is largely effected by the depth, as discussed in 3.2.3.) Since the real-world devices are not *set* within an infinite substrate, the differences in simulated response need to be considered, namely to make sure that the use of PMLs makes the results obtained valid.

Firstly the two boundaries, parallel to the x_3 axis, at either end of the simulation domain are considered. Clearly the propagating SAW emitted from each end of the transmitting IDT will be reflected off each of these edges, as shown in 4.20, causing some interference with the SAW between the two IDTs. As discussed in [12], this interference manifests itself as a high frequency oscillation over the top of the sinc response of the SAW device, which can be clearly seen in 4.14 of the response of a real-world SAW device. Modelling such an effect is by no means trivial since the boundary conditions required would not be simple, hardwall boundaries, but more like those implemented for the simulation of ridge structures in 5.1.2. Furthermore, these high frequency oscillations in the response do not significantly affect the operation of real SAW devices and therefore do not need to be included within the model.

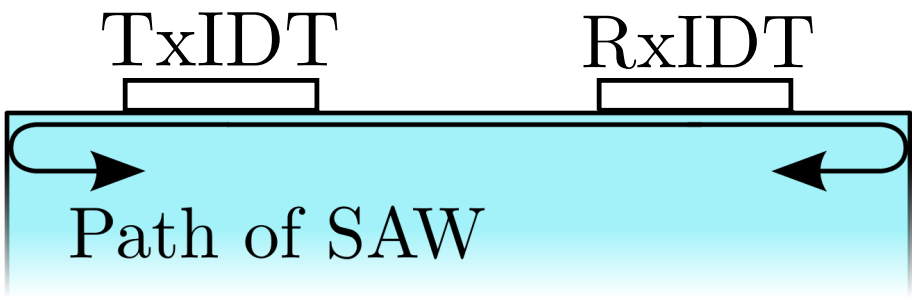

Figure 4.20: Path of SAW when it reflects off the side edges of a real-world SAW device.

The bottom edge of SAW devices play much less of a role in the response because they are normally of the order of 100 μm deep, meaning that almost none of the SAW energy will reach that depth. However, since IDTs don't purely excite SAW modes (as discussed in 4.3.3), there is the possibility of part of the acoustic energy bouncing off the bottom boundary, as shown in 4.21. Since SAW devices will physically be mounted onto some sort of holder modelling this problem would be more difficult, although this will actually act to dissipate much of the SAW energy hence why the effect of this bottom reflection is never seen. Therefore reflections off the bottom of the SAW device need not be considered.

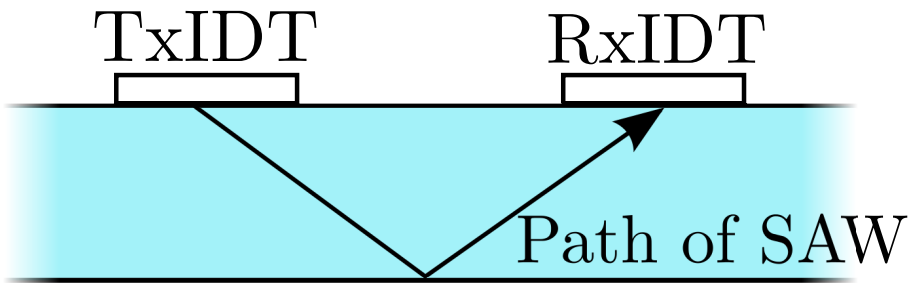

Figure 4.21: Path of SAW when it reflects off the bottom edge of a real-world SAW device.

4.3.5 Experimental validation of IDT model

As the model of SAW devices has been built up to a stage where it can model the signal on the receiving IDT, there is the possibility of comparing the results with experimentally measured values in order to give experimental validation to the model. As discussed in 4.3.4 however, the limitations of the model of receiving IDTs mean that the absolute amplitude of the signal is arbitrary and so cannot be used in an experimental comparison. Furthermore, since the acoustic wave model contains no loss mechanism, the comparison would be valid anyway. A property which can be used for the comparison however is the full-width half-maximum (FWHM) of the frequency response of the device. Not only is this a good property to compare because it is well documented experimentally that the FWHM decreases as the number of finger-pairs within each IDT is increased [12], but also because it does not depend on the amplitude of the response, only the dimensions of the of the IDTs and the form of the excited SAW (in terms of its frequency and speed of propagation). Such a comparison will then provide a very strong indication whether the form of the excited SAW is of the correct form.

For this comparison, a series of SAW devices were fabricated in the GaAs material system with a SAW propagation direction of [1,1,0] with a different number of finger-pairs in the IDTs on each device (fabricated by Dr. M. Salih). The number of finger-pairs was the same in each IDT on each device and varied though 10-, 20-, 40-, 60- and 80-finger-pairs. The separation between the two IDTs in the x_1 direction was 2.6 mm. The same set of devices were modelled using the FDTD simulation to produce a series of IDT frequency responses for each different number of finger-pairs. The FWHM of both the experimental and simulated responses was then taken and is shown figure 4.22.

Clearly the two are in very good agreement, giving a very strong indication that the SAW simulation is a viable method for modelling IDT devices. Furthermore, comparing the overall shape of the frequency response for the 40 finger-pair device in 4.23 both experiment and simulation are in good agreement, although the simulated results predict a higher received signal because there is no loss mechanism within the FDTD model.

4.3.6 Mass loading

The mass loading effect is something that effects IDTs that emit at a very short wavelength such that the width of the IDT fingers become comparable to their height. When this happens the weight of the metallisation of the IDT on the surface of the substrate begins to affect the

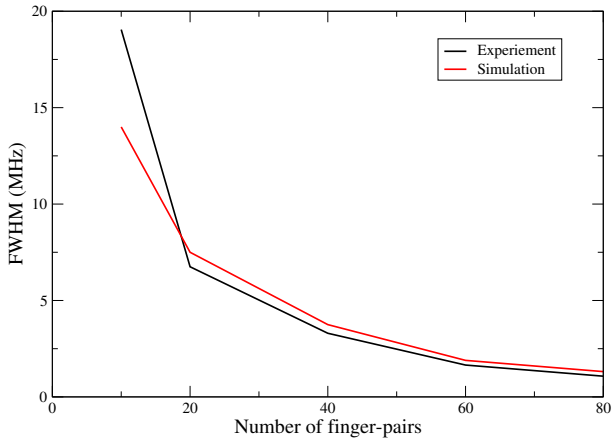

Figure 4.22: Comparison of the FWHM of the frequency response for SAW devices various numbers of finger-pairs found via experimental measurement and simulation.

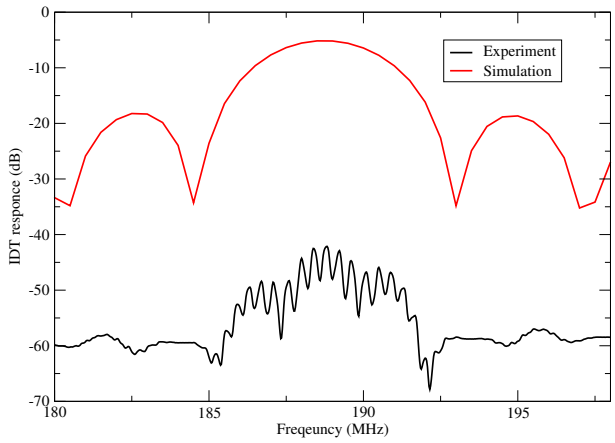

Figure 4.23: Comparison of the simulated and experimentally measured frequency response for a 40 finger-pair SAW device on GaAs with a [1,1,0] propagation direction.

mechanical motion of the propagating wave underneath the IDT. Considering the acoustic wave equation of motion

$$\rho \frac{\partial u_i}{\partial t} = \frac{\partial \sigma_{ij}}{\partial x_j},$$

clearly mass loading will effectively increase the density underneath each of the IDT fingers, since there will be more mass at those points to oscillate as the acoustic wave passes underneath. Furthermore, since the velocity of the acoustic wave can be approximated to

$$v_s \approx \sqrt{\frac{\mathsf{C}}{\rho}},$$

it can be assumed that the velocity of the SAW will slow down underneath each IDT therefore lowering the resonant frequency of the IDT device; something which is seen experimentally [13].

This approximation is crude however. In reality, a portion of the acoustic wave energy will enter the metallised region and effect the way the SAW propagates. However, even at very high SAW frequencies of $> 3\,\text{GHz}$ where the width of each finger is around 800 nm, the depth of metallisation is still only $50 - 150\,\text{nm}$, meaning that the fingers are still much wider than they are high implying that not too much of the acoustic wave will enter the metallised region.

In order to simulate the mass loading effect using the assumption that only the material density will change underneath each IDT finger, firstly that change in density must be defined in terms of the thickness of the IDT metallisation. This is defined as

$$\Delta\rho = \frac{\rho_{\text{IDT}} h}{\delta x_3} \tag{4.20}$$

where $\Delta\rho$ is the difference in density from mass loading, ρ_{IDT} is the density of the IDT metal and h is the height of the IDT metallisation. $\Delta\rho$ may then be added to ρ at all grid points within the FDTD simulation that are on the substrate surface and directly below any surface metallisation, i.e. the IDT fingers. Gold is normally used for IDT metallisation but for high frequency SAW devices aluminium is often used because of its lower density.

Figure 4.24 shows the simulated resonant frequencies found for a SAW device, with 10 finger-pair IDTs with a design wavelength of 840 nm, on GaAs ([1,1,0] propagating) for various metallisation depths with the IDT metal being both gold and aluminium. Clearly the frequency response of the device drops as the thickness of the surface metallisation increases. Furthermore, as expected, the frequency drops off quicker with gold IDTs since these have a much higher density than aluminium. What is also clear is that the ranges of surface metallisation thicknesses used here are much to small. This was done because increasing the thickness much beyond these ranges (i.e. $> 30\,\text{nm}$ for aluminium and $> 4\dot{\text{n}}\text{m}$ for gold) caused instabilities within the simulation because the discontinuity in the density was too great from the grid points underneath the IDT fingers to those not. This problem may very well be to do with the incorrect form of (4.20), and an empirical scaling factor could be added to compensate for this.

Examining the frequency responses of aluminium in 4.25 and gold in 4.26 shows that the amount of SAW energy captured by the receiving IDT increases as thicker metallisation is used. While this seems counterintuitive, it is a real effect and occurs because the increased mass of each finger gives each finger an increased momentum as it vibrates from generated SAW passing underneath it. This, in tern, increases the magnitude of the oscillations resulting in a larger acoustic wave [61].

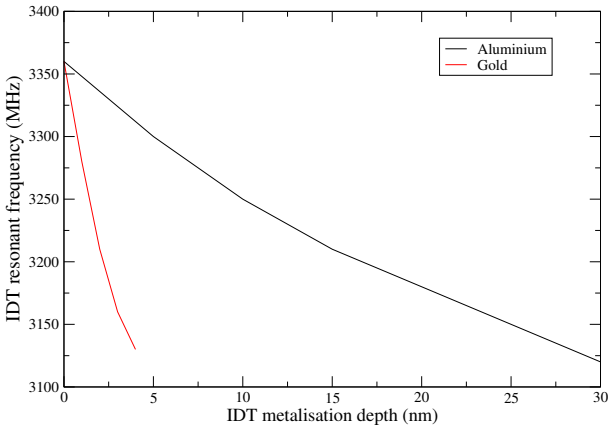

Figure 4.24: Simulated resonant frequency of the IDTs on a SAW device within the GaAs material system, [1,1,0] propagation direction, for gold and aluminium with different thickness of metallisation.

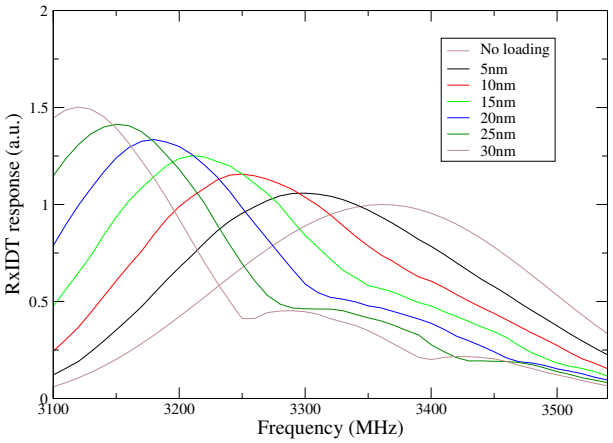

Figure 4.25: Normalised frequency responses for mass loaded SAW devices in the GaAs material system, [1,1,0] propagation direction, for aluminium metallisation with a range of thicknesses.

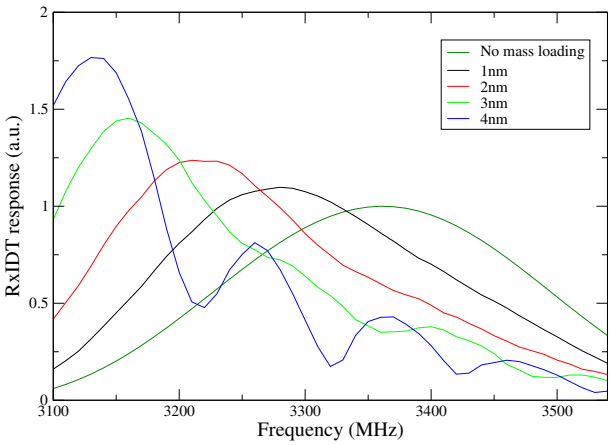

Figure 4.26: Normalised frequency responses for mass loaded SAW devices in the GaAs material system, [1,1,0] propagation direction, for gold metallisation with a range of thicknesses.

Chapter 5

Surface acoustic wave modulation of quantum cascade lasers

As quantum cascade lasers (QCLs) made out of III-V semiconductor heterostructures exhibit the piezoelectric effect, it is plausible that distributed feedback-like modulation of the laser cavity may be achieved via acoustic waves. This is achieved by applying a propagating acoustic wave through the laser cavity, which induces an electric field wave that redistributes free carriers within the laser active region (AR), causing a periodic modulation of the gain and refractive index. Unlike distributed feedback lasers where this modulation is caused by altering the physical structure of the device (normally by etching through the surface metallisation into the top contact layer of the QCL [26]), the periodicity of acoustically-induced modulation may be varied by as much as the device generating the acoustic wave allows, therefore allowing the output frequency of the laser selected via feedback to be tuned. This idea was first suggested by Kisin in 2003 [24].

There has been experimental work to fabricate a surface acoustic wave (SAW) modulated QCL device by Salih *et al.* which is summarised in [36]. Figure 5.1 shows the structure of the device which has been fabricated. This device consists of a QCL ridge, using the Luo *et al.* three-well AR [38], bonded to the top of a SAW device (consisting of a transmitting/receiving IDT pair on a GaAs substrate). The long axis of the QCL is aligned to the propagation direction of the SAW (which is the [1,1,0] crystal axis). A more detailed description of the fabrication and structure of this device may be found in [36].

As the SAW is generated within the substrate of this device, it relies upon the SAW energy moving from the substrate up into the QCL in order for modulation of the QCL carrier concentration to be achieved. Since placing the QCL ridge on top of the substrate

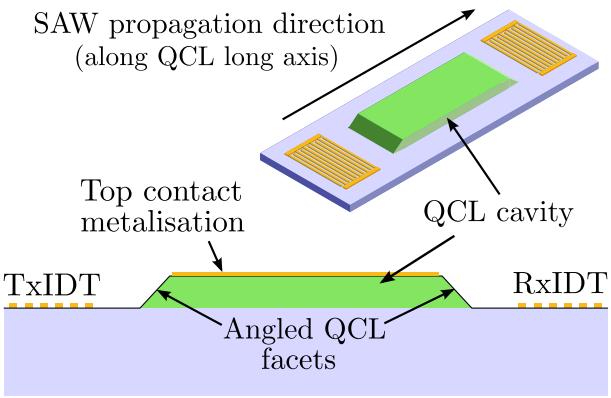

Figure 5.1: The SAW modulated QCL device design as proposed by Salih *et al.* [36] consisting of a QCL cavity mounted on top of SAW device. The propagation of direction of the SAW is aligned within the QCL long-axis and the QCL facets have been angled to try and help the SAW move from the substrate into the QCL cavity.

causes a discontinuity in the surface which the SAW travels along which is comparable to the SAW wavelength [1], it is likely that a significant portion of acoustic wave energy will be scattered at this discontinuity into a bulk mode which will not propagate through the QCL ridge (and therefore not contribute to modulating the QCL), as illustrated in figure 5.2. Therefore this device has two structural features which try to minimise the scattering of the SAW at this discontinuity. Firstly, the thickness of the QCL AR has been reduced from a typical thickness of $\sim 10\,\mu$m to $5\,\mu$m in order to reduce the discontinuity on the surface [39]. Secondly the facets of the QCL ridge are wet etched to produce a sloped surface, which acts as a taper, removing the sudden step from the substrate surface to the QCL ridge surface. Whilst altering the angle of the facets will reduce the effectiveness of the QCL ridge as a resonant cavity (as well as affecting the beam pattern emitted by the QCL), QCL ridges with sloped facets have been show to lase by Salih *et al.* [36]. Despite these features designed to enhance the modulation of the QCL carrier concentration by the SAW, no observable modulation to the QCL emission has been observed from this device. Furthermore, there has been no investigation into how much of the SAW energy moves from the substrate into the QCL AR and therefore whether this device represents a feasible design from which SAW modulation of the QCL AR could be achieved.

[1] The SAW wavelength must be approximately double the wavelength of light at the emission frequency within the QCL ridge in order to achieve distributed feedback around the lasing frequency [26], which for the Luo *et al.* AR is $\sim 12.5\,\mu$m; where as the QCL ridge height will be between 5 and 15 μm as discussed above.

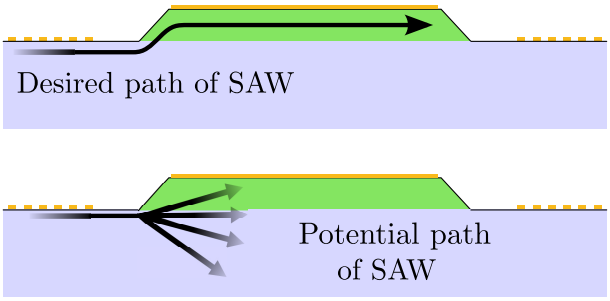

Figure 5.2: The potential paths of the SAW as it moves from the substrate into the QCL cavity, either moving completely up into the QCL cavity and staying as surface bound propagation mode *(top)* or scattering in to a bulk propagation mode as it hits the discontinuity in the surface such that much of the energy of the acoustic wave does not enter the QCL cavity and does not contribute to modulation *(bottom)*.

Within this chapter, a full model of SAW propagation through a QCL device is developed by extending the FDTD acoustic wave model of SAW devices presented in chapter 4. A self-consistent rate equation model [43, 44, 45, 41] is discussed next which, using the modulation in carrier concentration calculated by the model of SAW propagation through QCLs, is used to simulate the induced change in complex refractive index. The change in complex refractive index may then be used determine the strength of the distributed feedback by utilising coupled wave theory [62]. Finally, using the results from the analysis within this chapter a recommended design for a SAW modulated QCL is presented, designed to maximise the modulation of the carrier concentration within the QCL AR by the SAW and therefore the strength of the distributed feedback.

5.1 Extensions to the SAW device model

As the model of SAW devices is required to simulate SAW-modulated quantum-cascade-laser device, there are several extension which need to be made to the model before it may be used to simulate these, or any other more complicated devices involving SAW propagation. These will be discussed here.

5.1.1 Surface metallisation

Introducing surface metallisation with an applied bias into the model is relatively simple because of it's similarity to including IDTs. Since Poisson's equation is already solved at

every time step, and the Poisson solver includes the functionality to fix the potential at any points within the simulation, the problem simply becomes defining the position of the metallised region as well as the voltage applied to it (including any time variation in that voltage). For a free metallised surface, not connected to any voltage, the method is identical to that used for modelling 2DEG's only placing the conductive region on the surface, therefore it will not be discussed here.

Fixing the potential at the surface of the simulated structure pins the SAW induced potential to that voltage directly underneath the metallised region. This *pinning* effect can be seen in 5.3 where a small region in between the transmitting and receiving IDTs has been fixed to various potentials on a GaAs substrate.

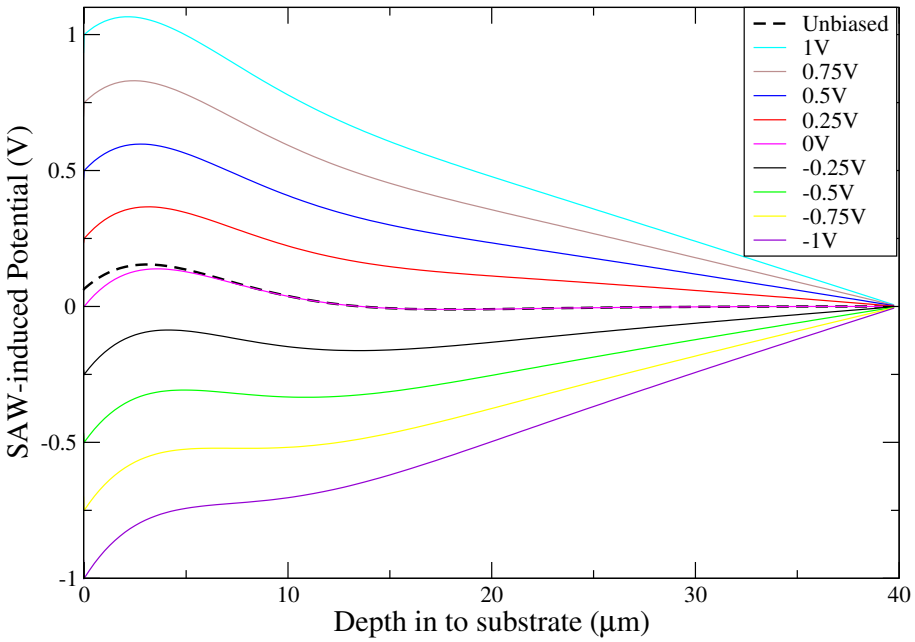

Figure 5.3: The profile of the SAW induced potential underneath a metallised region on the surface connected to voltages ranging from ±1 V (including the unbiased case), with a SAW wavelength of 15 μm on [1,1,0] GaAs.

5.1.2 Ridge structures

In order to simulate the interaction between the SAW and QCL ridge which is placed in its path, the surface boundary condition of the acoustic wave simulation must be altered to allow ridge structures can be defined. This problem is far from trivial since the sides of the QCL ridge do not have a regular shape, as shown in [37], because of the wet etching process used

to create the ridge structure. This makes the edges of the device slightly curved, such that the gradient of the edge of the ridge increases towards to the top of the ridge. Furthermore, because of the square nature of the grid used in the acoustic wave simulation, any surface shape that runs at an angle not parallel or perpendicular to the x_1 and x_3 will be subject to some quantisation (although this will clearly become less of an issue when the ridge height is much larger than the SAW wavelength since the grid steps, δx_1 and δx_3, are dependant upon the wavelength).

Therefore as a first approximation, only *square* ridges will be considered, such that their edges are either parallel or perpendicular to the axes x_1 and x_3.This approximation makes defining the surface boundary for the flat areas along the ridge easy, since they are identical to the regular surface boundary, that is the differential of the stress across the surface must be zero, i.e.

$$\left.\frac{\partial \sigma_{i3}}{\partial x_3}\right|_{x_3=\text{surface}} = 0 \qquad \text{for } i = 1, 2, 3, \tag{5.1}$$

for horizontal surfaces, and

$$\left.\frac{\partial \sigma_{i1}}{\partial x_1}\right|_{x_1=\text{surface}} = 0 \qquad \text{for } i = 1, 2, 3, \tag{5.2}$$

for vertical surfaces, where *surface* defines any point in space which lies on a surface. However, defining the boundary conditions for the corner points is not straightforward. For the *concave* corner, the horizontal and vertical boundary conditions will combine to give the bulk equations of motion, i.e. neither (5.1) or (5.2) apply. For the *convex* corner on the other hand, the two surface boundaries combine to set the left-hand-side of the equations of motion zero (since the $\frac{\partial}{\partial x_2}$ terms are already zero due to the invariance in x_2), i.e. both (5.1) and (5.2) apply. Clearly neither of these corner conditions are correct since the displacement at the convex corner cannot be zero without some external force pinning it there, and bulk conditions cannot be used for concave corners since there are points *missing* from outside the corner which are required to solve the equations of motion. Therefore, a further approximation must be applied to find suitable boundary conditions for corners along the surface.

The displacement in the spatial dimensions for the concave corners is calculated as the average of the displacement at the four adjacent points, i.e. for a concave corner at the point i, j

$$u_i(i,j) = \frac{u_i(i-1,j) + u_i(i+1,j) + u_i(i,j-1) + u_i(i,j+1)}{4} \qquad \text{for } i = 1, 2, 3. \tag{5.3}$$

The rationale behind this is that the difference in the displacement should be small enough to approximately be linear between points.

For the convex corners, where the corner point should be free to move under any applied force, but not move in a way which adds any energy into the system, the assumption is made that the distance between the corner point and the two adjacent points on the surface is constant, and of the size of the spatial grid step, as shown in 5.4. This is done only for the displacements u_1 and u_3 to simplify the trigonometry, since only ridge structures in the GaAs material system with [1,1,0] propagating SAWs are considered within this thesis (and the u_2 displacement is always zero in this material system). However, this method could be extended to include u_2 displacement. In order to find the displacement at the corner point i, j, firstly the sides of the dashed triangle, a and b are found by

$$\begin{aligned} a &= |\delta x_1 - u_1(i, j+1) + u_1(i+1, j)|\,, \\ b &= |\delta x_3 - u_3(i+1, j) + u_3(i, j+1)|\,, \end{aligned} \tag{5.4}$$

with the hypotenuse of the dashed triangle being given simply by $\ell = \sqrt{a^2 + b^2}$. The angles η and φ may then be found using the laws of tangents and cosines,

$$\begin{aligned} \tan\eta &= \frac{b}{a}, \\ \cos\varphi &= \frac{\ell + (\delta x_1)^2 - (\delta x_3)^2}{2\ell\delta x_1}. \end{aligned} \tag{5.5}$$

Finally, the two components of displacement at i, j may be found by

$$\begin{aligned} u_1(i, j) &= u_1(i+1, j) - \delta x_1 \sin(\eta + \varphi - \frac{\pi}{2}), \\ u_3(i, j) &= u_3(i+1, j) + \delta x_1(\sin(\eta + \varphi) - 1). \end{aligned} \tag{5.6}$$

This analysis has been done for *left-hand* convex corners, i.e. where the positive direction of the axes is as shown in 5.4, however the same analysis is readily applied to *right-hand* convex corners where the positive x_1 axis points in the opposite direction.

These corner boundary conditions were applied to a simple ridge structure in the [1,1,0] propagation direction for GaAs material system, with a SAW generated by a 10-finger-pair IDT with $\lambda = 15\,\mu$m, with varying ridge heights. The u_1 displacements after 40 ns, shortly after the acoustic wave has reached the ridge, is shown in 5.5 and 5.6. These show that the magnitude of the acoustic wave appearing after the ridge seems to increase as the ridge height increases, implying that ridge structure boundary conditions are adding energy into the system. Furthermore, the correlation between ridge height and magnitude seems to imply that the energy is being added to the system via vertical surface boundary, however because this is identical to the horizontal surface boundary (only rotate via 90°) which works well this is not likely. What is more likely is that the concave corner boundary is adding energy

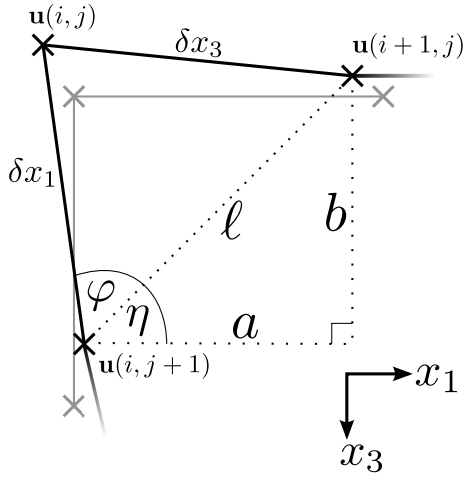

Figure 5.4: Labeling of the angles and lengths involved in calculating the displacement at a convex corner point i, j at the surface. The grey crosses represent the positions of the grid points, each separated by either δx_1 or δx_3, and the black crosses represent the actual positions of those points at some moment in time, with the displacement, $\mathbf{u}$, being the distance between the two.

to the system which travels up the vertical surface boundary before being reflected back by the convex corner boundary. If the ridge height is small then this *added* acoustic energy is reflected and absorbed by the bottom PML before it passes by too many bulk grid points. For larger ridge height this acoustic energy will pass by more bulk crystal (i.e. within the ridge) before it is reflected off the top of the ridge and absorbed by the PML. This appears to be consistent with the shape of the acoustic wave next to the start of the ridge in figures 5.5 and 5.6.

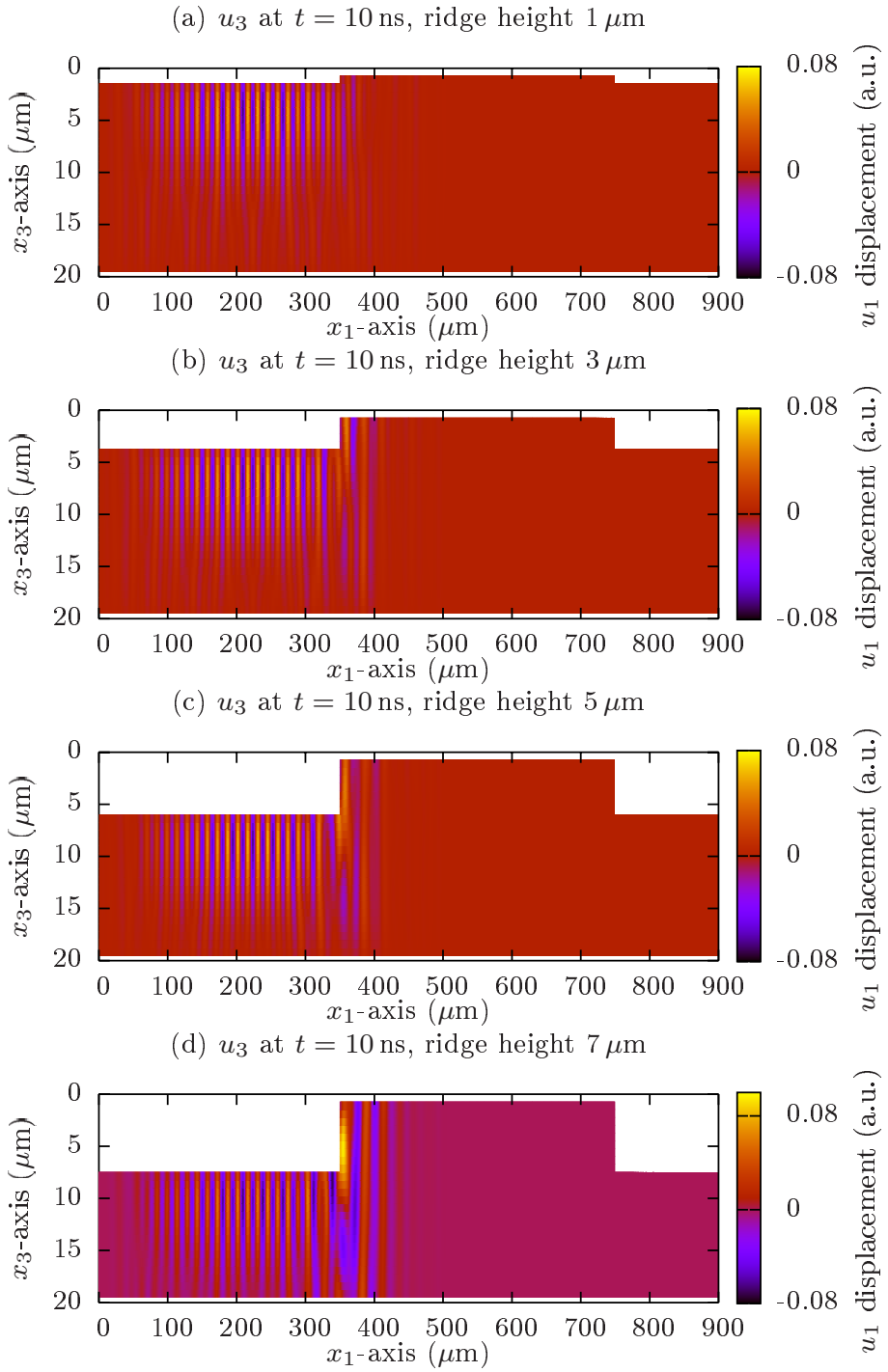

Figure 5.5: Propagating SAW (u_1 displacement) from a 10-finger-pair IDT on [1,1,0] ($\lambda = 15\,\mu\mathrm{m}$) propagating GaAs after 40 ns, propagating in ridge structures of varying heights, between 1 and 7 μm.

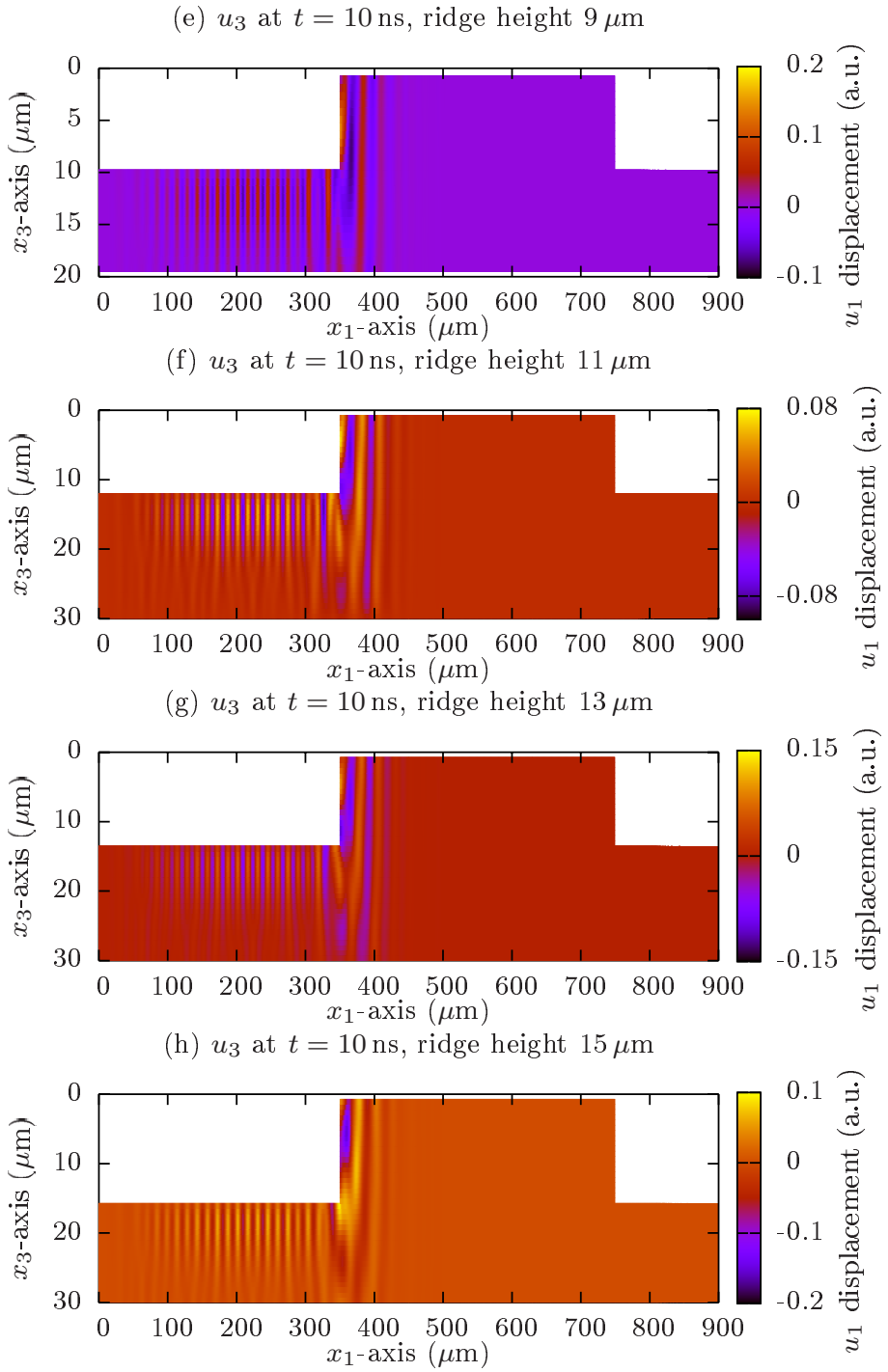

Figure 5.6: Propagating SAW (u_1 displacement) from a 10-finger-pair IDT on [1,1,0] ($\lambda = 15\,\mu\mathrm{m}$) propagating GaAs after 40 ns, propagating in ridge structures of varying heights, between 9 and 15 μm.

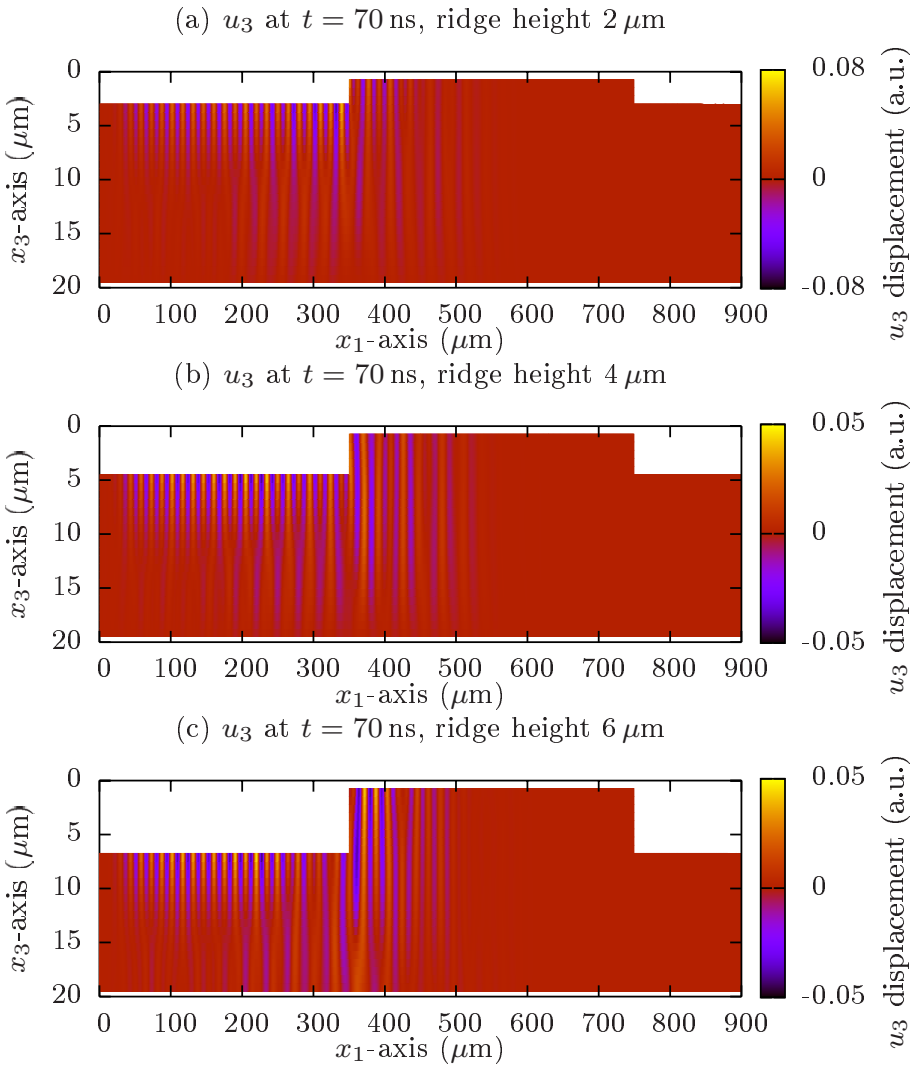

Figure 5.7: Propagating SAW (u_3 displacement) from a 10-finger-pair IDT on [1,1,0] ($\lambda = 15\,\mu$m) propagating GaAs after 70 ns, propagating in ridge structures of varying heights, between 2 and 6 μm.

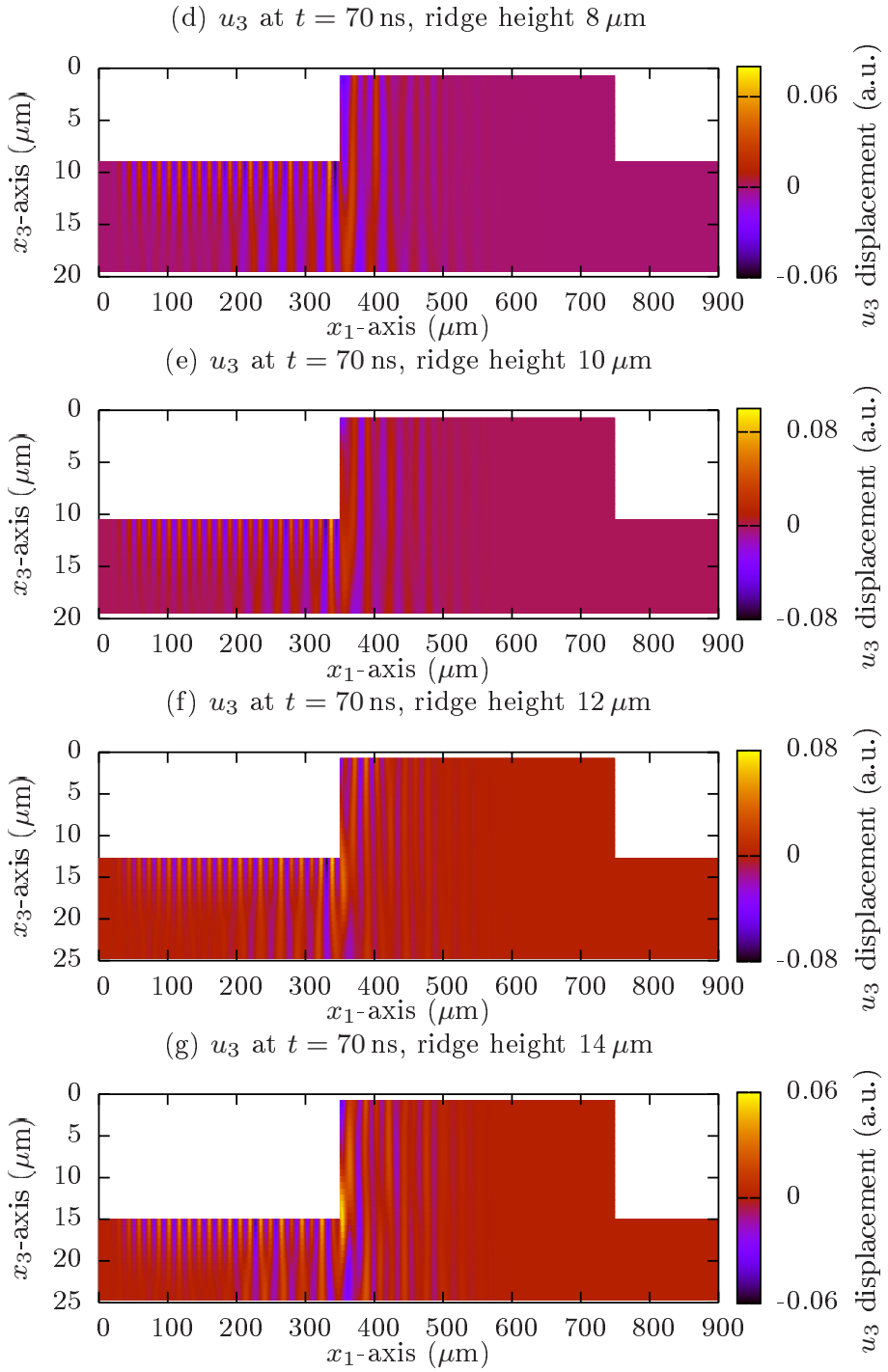

Figure 5.8: Propagating SAW (u_3 displacement) from a 10-finger-pair IDT on [1,1,0] ($\lambda = 15\,\mu$m) propagating GaAs after 70 ns, propagating in ridge structures of varying heights, between 8 and 14 μm.

Alternative ridge boundary conditions

While the ridge boundary conditions derived above are not perfect, they are the most stable of all the different types tried within the acoustic wave simulation to date. Some of the other boundary types that have been tried include

Forward/backward differencing

This tries to remove the need for the missing point in the bulk approximation to a concave corner by using forward/backward finite differencing in the differentials for the displacement, **u**. This boundary was also applied to convex corners without success. As applying a forward/backwards difference introduces a larger computational error within the interlaced grid scheme, this approach was also tried in a simulation where the auxiliary variable **V** had not been introduced to the equations of motion (and hence without PMLs). While this reduced the computational load of this approach, it still led to instabilities and was unsuccessful.

Constant differential along boundaries

The value of the first and second order differential was fixed assumed constant around each corner to try and approximate the displacement at the corner points. This method was also unstable.

Masses-on-springs derivation of surface boundary

A masses-on-springs view of the discretisation of the acoustic wave equations of motion was used to derive the surface boundary condition, and it was found that this altered the equations of motion at a horizontal surface from

$$\rho \frac{\partial u_i}{\partial t^2} = \frac{\partial \sigma_{i1}}{\partial x_1},$$

to

$$\rho \frac{\partial u_i}{\partial t^2} = \frac{\partial \sigma_{i1}}{\partial x_1} + \sigma_{i3}.$$

From this approach, the combination of horizontal and vertical boundaries at the corners of a ridge do not become zero or bulk, but rather corner boundary conditions drop out of the bottom. This approach seemed promising, but is incompatible with the IDT excitation method used as it leads to instabilities at the points of fixed potential within the simulation domain. It is unclear whether these instabilities come from the new surface boundaries or from the assumptions made in the IDT model.

Analysis of ridge boundary conditions

All of the methods used to derive ridge boundary conditions for the FDTD acoustic wave model have been applied to a ridge with a vertical surface and right-angled corners which, as discussed previously, represents the worst-case in terms of scattering of an incident acoustic wave. Since all of these methods resulted in boundary conditions which were unstable (i.e. they added energy to acoustic simulation domain), this implies that at at least one of the right-angled corners the magnitude of one of the differential terms in the equations of motion is becoming unrealistically large, producing an acoustic displacement which is larger than physically possible. However as the right-angled corners themselves are physically realistic, it is inferred that the instabilities introduced to the acoustic wave model by the addition of ridge structure boundaries are not a problem with the boundary conditions, but a fundamental problem with trying to simulate an unrealistic structure. Therefore in order to simulate realistic ridge structures which have curved corners with a continuum model that utilises a rectangular discretisation grid (such as the FDTD acoustic wave model), a scheme for deriving boundary conditions for curved boundaries which lie in between grid points would have to be used; analogous to that used in [63] for solving Maxwell's equations around curved waveguides — This is set as further work in chapter 6.

While the instabilities introduced by the ridge structure boundary conditions stop any quantitative analysis of their results, a qualitative discussion about the fact that these instabilities occur can still give insight into the scattering of acoustic wave energy by a physically realistic ridge structure. The fact that right-angled corners cause at least one of the differentials within the equations of motion to be large enough to cause instabilities implies that for a physically realistic ridge structure (i.e. with curved corners) these same differentials would still be large, but not large enough to cause instabilities, and therefore there would be a significant amount of scattering from a surface bound mode to a bulk mode. While the SAW-QCL structure from [36] does employ sloped facets on the QCL ridge to try to minimise the this acoustic wave scattering, the length of this sloped section along the propagation direction is of the order of the SAW wavelength, implying that there will still be significant scattering. Furthermore, although the height of the QCL ridge has been reduced to a minimum of 5 μm, the instabilities in the ridge structure simulations appear for ridge heights of 5 μm and above, implying that there will still be significant scattering.

Although these two structural features will help reducing this scattering to some degree, the line of reasoning followed above implies that they will not be enough to prevent significant

scattering of the surface mode into a bulk mode of acoustic wave propagation that will not contribute to the modulation of the QCL active region.

5.1.3 Buried two-dimensional electron gas

The quantum wells within the QCL have a high concentration of free carriers, which are free to move in-plane of the quantum well structure, i.e. in the x_1 direction within the FDTD simulation. These free carriers will move under the influence of the SAW-induced piezoelectric potential, causing areas high and low carrier concentration along the x_1 axis, and therefore a modulation in the QCL active region. However, the potential from the free charge will act to screen the potential induced by the SAW meaning that, depending on the thickness of QCL active region, the SAW potential may not influence the carrier concentration through the entire height of the QCL.

Since the SAW simulation domain is effectively two-dimensional, at its most complicated, modelling this problem involves finding the charge density at every point in time for an area of free charge within the simulation domain. This, however, is not easy as it requires the confinement of a two-dimensional charge-well to be considered, and the problem is simplified greatly by assuming the free charge area is made up of a series of layers of two-dimensional electron gas in the x_1-x_2 plane. This approximation is not too far from reality since the quantum well structure within the QCL does provide confinement in one direction, and although in reality free carriers do tunnel from one quantum well to another the carrier concentration within one well should remain reasonably constant as the device is electrically pumped. This reduces the problem to finding the one dimensional carrier concentration within a series of layers, however the confinement of one-dimensional potential must still be considered. While a classical approach could be to find the evolution of carrier concentration over time, hence ignoring electrons tunnelling from one area of confinement to another, the problem may be simplified further by realising that only a steady state solution is required to determine the strength of the modulation of the carrier concentration. Furthermore, since (to a first approximation) carriers will be injected into the QCL active region evenly over the area of the top of the device, finding the time varying carrier concentration for a layer of 2DEGs that does not take this into account would not valid. Therefore the charge within each 2DEG can be assumed to be spread out evenly at every time step, and modulated accounting the SAW potential at that time step.

Firstly, examining the dielectric relaxation time [64],

$$\tau = \frac{\varepsilon}{\sigma},$$

where ε is the permittivity and σ is the conductivity, for GaAs with a relatively low carrier concentration of $\sim 10^{15}\,\mathrm{cm}^{-3}$ gives a time of the order of 10 ps, much shorter than the nanosecond timescales of the SAW propagation, meaning that free carriers will respond adiabatically to the SAW potential. Secondly, it is assumed that all the donor atoms within the structure will donate an electron, as is done within the rate-equation model of QCLs, meaning that the number of electrons within each 2DEG is equal to the number donors within that volume, $\mathrm{N_D}$. Finally, assuming that the charge displacement induced by SAW will never be large enough to fully deplete any area within the 2DEG of carriers, then the electron charge within the 2DEG, ρ_e, is given by

$$\rho_e = -q\mathrm{N_D} - \rho_{\mathrm{SAW}}, \tag{5.7}$$

where $q\mathrm{N_D}$ is the charge from the donors and ρ_{SAW} is the charge induced form the SAW. While it may seem like a problem that this assumption is only valid when no area of the 2DEG is fully depleted of carriers, since the purpose of modeling the QCL active region is to determine the strength of the modulation and in a steady state when the carriers become depleted this is effectively the maximum modulation, this assumption still allows this purpose to be fulfilled.

As the simulation of ridge structures within the SAW model appears to add energy into the system, simulating a buried 2DEG within a ridge structure would not give valid results. As an interim solution, a 2DEG may be placed within the substrate itself with no ridge structure on the surface. Clearly underneath the transmitting IDT there will be a much higher potential due to the IDT potential itself, which will induce a higher modulation of the carrier concentration. Therefore the modulation of the carrier concentration underneath this IDT can be ignored.

Figure 5.9 shows the modulation of the carrier concentration of a 10 μm thick layer of 2DEG with various doping densities, buried 5 μm below the surface of the channel of a SAW device on a [1,1,0] GaAs substrate. The doping density, N_d, and therefore the unperturbed carrier concentration, varies from $1\times10^{17}\,\mathrm{cm}^{-3}$, i.e. the carrier concentration in the three-well active region QCL by Luo *et al.* [38]. The voltage on each finger of the transmitting IDT was set 0.5 V. Clearly, even with this relatively small value for the voltage on each finger of the transmitting IDT, at the carrier concentration of $1\times10^{17}\,\mathrm{cm}^{-3}$, the SAW induced charge

displacement is much higher than required to completely deplete regions of the 2DEG of carriers. In fact, the carrier concentration may be increased by an order of magnitude to $10\times10^{17}\,\mathrm{cm}^{-3}$ before the carrier concentration is too high for the SAW to fully deplete areas of carriers. Furthermore, the strength of the modulation does not appear to diminish into depth of the substrate indicating that the 2DEG does not screen the modulating effect of SAW. (The strength of the modulation does reduce towards the bottom 2DEG layers however this corresponds to the reduced SAW power at a depth of $15\,\mu$m, i.e. $1\,\lambda$.) It is thought that this is due to the piezoelectric effect being relatively weak within the GaAs material system, since the only mechanism by which the 2DEG layer may screen the SAW is through altering the induced SAW potential, which will feedback and alter the mechanical motion through the piezoelectric term within Hooke's law. In GaAs this feedback is relatively weak as the values within the piezoelectric tensor are small compared to, say, lithium niobate of quarts substrates (see appendix C for these material constants). Therefore the screening effect of the 2DEG layer will therefore not alter the mechanical motion of the SAW to a large degree, meaning that modulation of the carrier density will not change a great deal into the depth of the material.

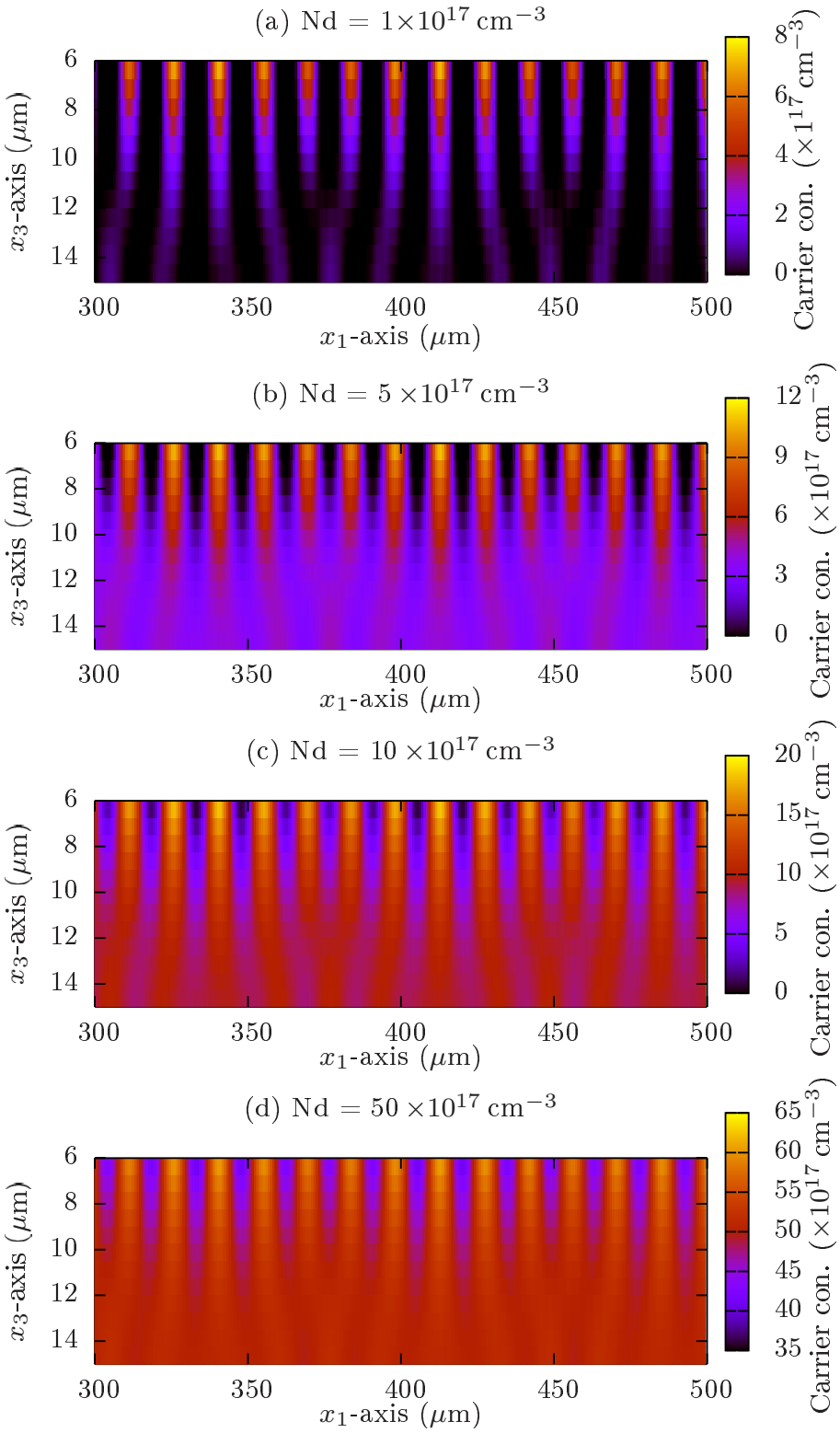

Figure 5.9: Modulation of the carrier concentration within a 10 μm thick layer of 2DEGs with various doping densities, buried 5 μm below the surface of the channel of a SAW device in the [1,1,0] GaAs material system. The voltage on each of the fingers of the transmitting IDT was set to 0.5 V and the doping densities, and therefore the unmodulated carrier concentrations vary from 1×10^{17} cm^{-3} to 50×10^{17} cm^{-3}.

To investigate how a change in SAW power effects the modulation in carrier concentration, the same structure was simulated and the voltage on the transmitting IDT was varied. The unmodulated carrier concentration was set to the carrier concentration in the three-well active region QCL by Luo *et al.*, $8.1 \times 10^{14}\,\mathrm{cm}^{-3}$. Figure 5.10 shows the results for several different applied voltages to the transmitting IDT. For this carrier concentration, the conductive layer is fully depleted of carriers with the voltage on the IDT fingers being lower 1 mV and even with the voltage as low 0.1 mV there is still a significant modulation in the carrier concentration.

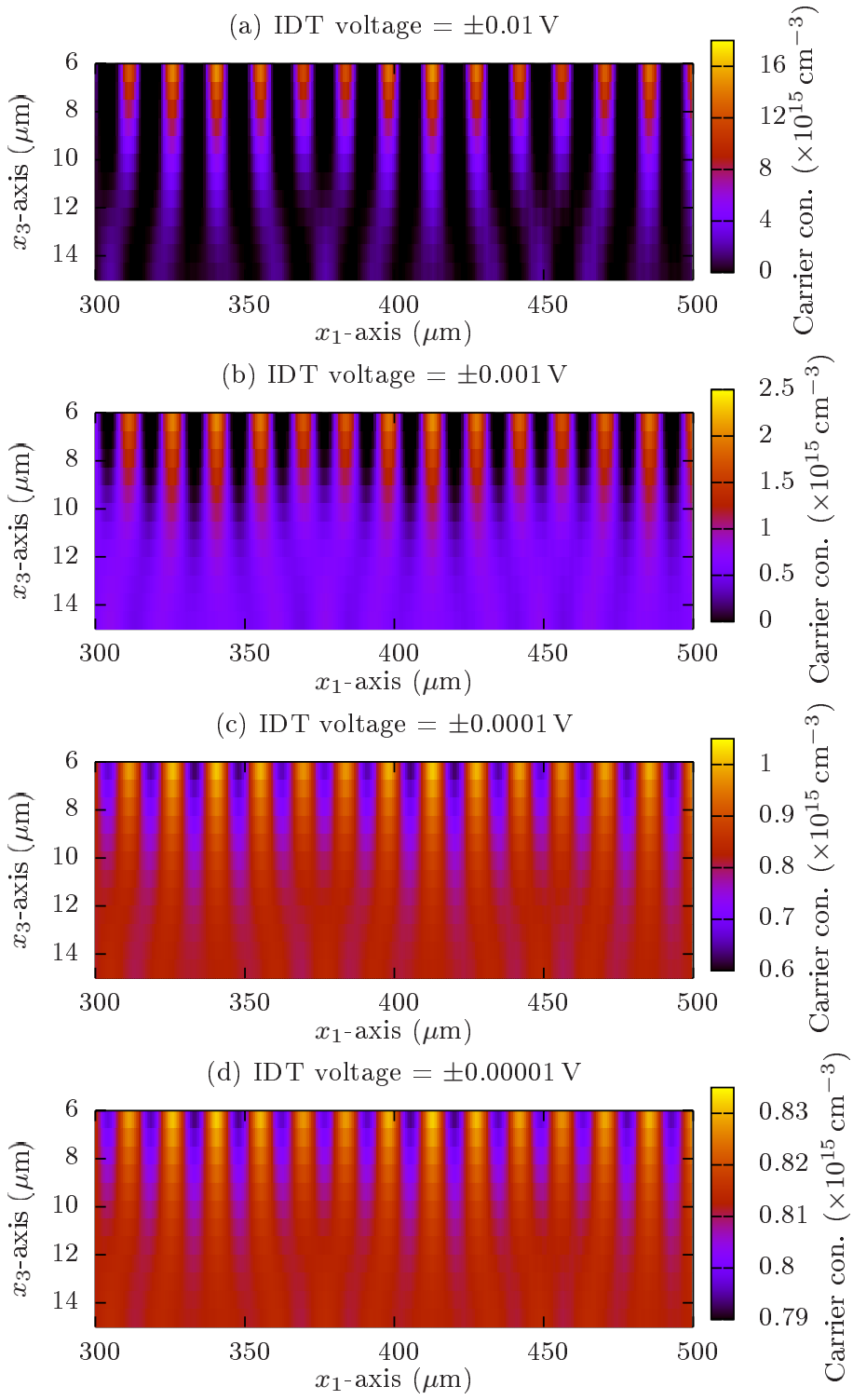

Figure 5.10: Modulation of the carrier concentration within a 10 μm thick layer of 2DEGs buried 5 μm below the surface of the channel of a SAW device in the [1,1,0] GaAs material system, with various voltages applied to the IDT fingers. The unmodulated carrier concentration was set to that of three-well active region QCL by Luo *et al.* [38], 8.1×10^{14} cm^{-3}.

These results indicate that the SAW does have more than enough power to fully deplete areas within the active region of a QCL of free carrier, therefore achieving the *maximum* amount of modulation. Since in real-world SAW devices the applied voltage to the contacts of the IDTs is of the order of several volts, even if the majority of the applied voltage does not reach the IDT fingers due to the mismatch in impedance from the contacts to the IDT bond-pads such that the voltage on each finger was say 1 mV, the QCL ridge would need to scatter more than 90% of the SAW energy into a bulk mode to stop a significant modulation in carrier concentration.

5.2 Rate equation model of quantum cascade lasers

The self-consistent Schrödinger-Poisson solution provides an accurate method for determining the electronic of a semiconductor heterostructure. For example, when simulating the three-well active region QCL device by Luo *et al.* the lasing transition is predicted to be 13.8 meV, corresponding to an emission frequency of 3.3 THz compared to the experimentally measured value of 3.4 THz. However, by examining the scattering rates between each state found within the heterostructure the model may be improved and extended to calculate other meaningful values which may be experimentally measured such as the frequency dependent gain or current-voltage curve of a QCL.

5.2.1 Structure of the rate equation model

The rate equation model is built on the structure of the self-consistent Schrödinger-Poisson loop, as in 2.3, however it includes the calculation of the scattering rates for each transition which, by solving the rate equations, allow the populations of each state to be found. The scattering rate calculations are all based upon Fermi's golden rule [65], which states that the lifetime of a charge carrier in some initial state $\langle i|$, which is subject to some time-dependent perturbation $\tilde{\mathcal{H}}$ that may cause it to scatter to some final state $|j\rangle$, is given as

$$\frac{1}{\tau_{ij}} = \frac{2\pi}{\hbar} \sum_{f} \left| \langle j|\tilde{\mathcal{H}}|i\rangle \right|^2 \delta(E_j - E_i) \tag{5.8}$$

where τ_{ij} is the average lifetime of a carrier in state $|i\rangle$ before it scatters to state $|j\rangle$ and the term $\delta(E_j - E_i)$, where E is the energy of each state, ensures that that the transitions is energetically possible. The form of $\tilde{\mathcal{H}}$ changes depending on the scattering mechanism which is being considered. These include alloy disorder scattering, acoustic phonon scattering, carrier-carrier scattering, carrier-photon scattering, interface roughness scattering, ionised

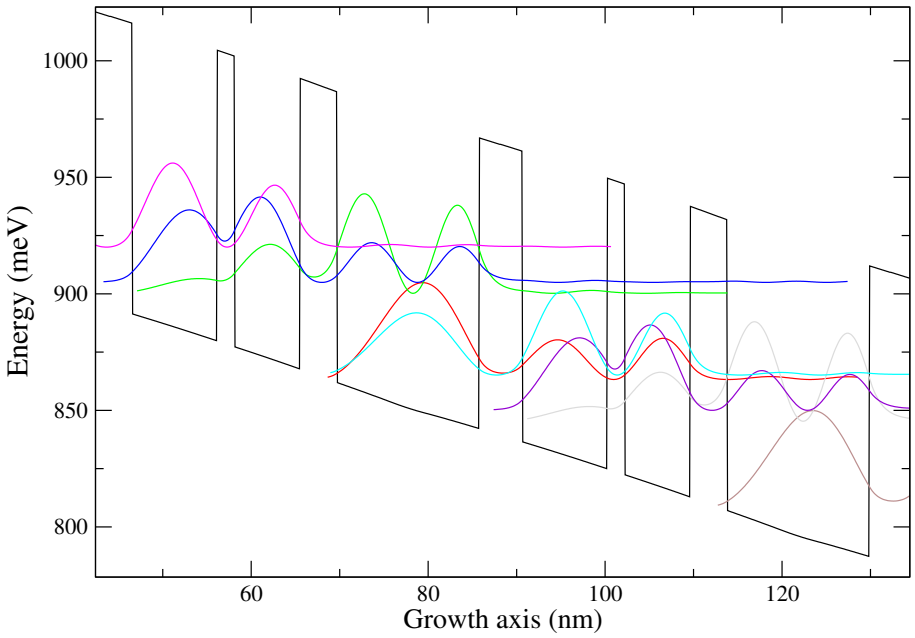

Figure 5.11: Wavefunction diagram showing two periods of the three-well active region structure by Luo *et al.* [38]

impurity scattering and longitudinal optical phonon scattering, full details of which can be found in [41]. If the system has more then two states, as in a QCL, then the total scattering rate for each state will be the sum the scattering rates for each transition from one initial state, i.e.

$$\frac{1}{\tau_i} = \sum_j \frac{1}{\tau_{ij}}$$

Once the total scattering rate of each state has been calculated, the rate equations may be solved. These come from the fact a carrier in one state may scatter in any other state within the system, as shown in 5.12, and therefore the total change in the population of any state over time is simply sum of emission and abortion rates for that state. This gives enough information to calculate the ratio of the populations of all states which, assuming that the total population is equal to the concentration of donor atoms, is enough to calculate the population of each state [41].

Summing the total scattering rates, both scattering out of and in to state $|i\rangle$, multiplied by the initial state population, n, for each gives the rate equations [41],

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = \sum_j \frac{n_j}{\tau_{ji}} - \left(\sum_j \frac{n_i}{\tau_{ij}} \right). \tag{5.9}$$

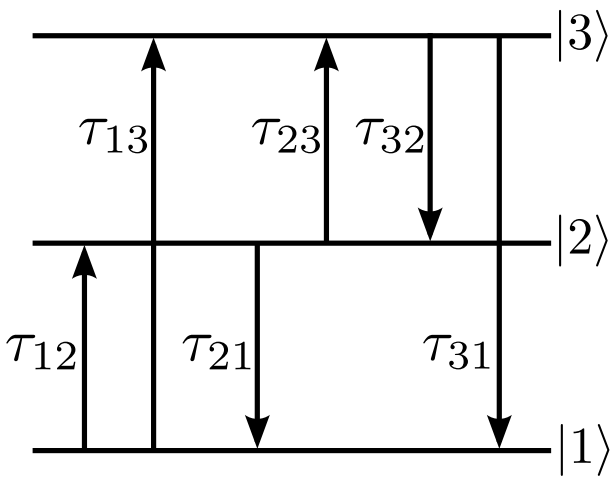

Figure 5.12: Scattering transitions within a three level system showing lifetimes of each transition.

Assuming the system is in a steady state, i.e. $\frac{\mathrm{d}n_i}{\mathrm{d}t} = 0$, the rate equations give a set of simultaneous equations which may be solved to find the ratio of the population in each state.

5.2.2 Object orientation of rate equation model

Due to the complexity and large number of varying tasks within the rate equation model, there is a considerable amount of computer code required to implement it. For ease of implementation, it is normal for each task to be broken into individual programs which are tied together by some master script, whose task is to execute all of the programs in the correct order, with the correct inputs, in order to run the simulation properly. For the specific rate equation model used in this thesis, the simulation has historically been developed on a Linux platform using Fortran and C program tied together by a Bash script, with communication between each program being done by reading and writing to and from files on the hard drive.

While this strategy for solving a complex problem such as the rate equation model does work incredibly well, moving to an object orientated implementation of each program introduces many advantages which are difficult (if not impossible) to achieve with a *linear* programing language. Aside from the obvious advantages such as the robustness of utilising objects under different conditions and the easy removal of duplicate code, as well as many other very well documented advantages (see any introductory book on object orientated programming, e.g. [66]), object orientation does a lot to help simplify complicated, computational code such as that in the rate equation model since simpler tasks can be done by a

relevant object rather than inside a calculation which may involve that object. Furthermore, the problem of the rate equation model is well suited to object orientation since so many of its components may be converted into objects. For example the form of the heterostructure may be stored in an *heterostructure object* containing the conduction band offset, effective mass profiles, etc., while the wavefunction and energy of each solution to the Schrödinger equation may be stored in *sub-band objects*, with a different object for each solution, which also contains information such as the population of the sub-band and Fermi energy along with methods for calculating the sub-band energy at a particular wavevector or the sub-band population up to some wavevector. Pairs of sub-band objects may then be stored within a *transition object*, which contains the scattering rate for the transitions but also methods for calculating the dipole matrix element between each sub-band and the population difference between sub-bands at some wavevector.

Due to the attractiveness of this implementation, a lot of work has been been put into the object orientation of the rate equation model, in conjunction with Dr. A. Valavanis (who has led the project) and A. Grier. This has involved refactoring much of the original rate equation model [43, 44, 45, 41] into a series of object orientated programs. Currently these programs are still tied together via a Bash script, however with much of the functionality being placed within a library, along with the object orientation, there is the prospect of being able to control each program from its own containing object and eventually tie the entire simulation together from within a single compiled program. This would remove the need for writing intermediate data to and from files since it could easily be passed from one object to the next, therefore improving the speed of the simulation. Furthermore, a library-based software solution allows for the user interface to be very easily improved with the possibility of a graphical front-end being added to control the simulation.

5.2.3 Complex refractive index calculation

Since it is the change in complex refractive index (i.e. the real component of refractive index and gain) that are modulated by the change in carrier concentration, these values need to be calculated. [25] presented a method for calculating the contribution to atomic susceptibility from one transition by using the dipole matrix element and population difference between the two states, from which the complex refractive index may be calculated. As these values come out from the rate equation model, this method can be implemented after the rate equation model has finished to calculate the complex refractive index.

From [25], the atomic susceptibility for a transition between two states is defined as

$$\chi = \chi^{'} - \mathrm{i}\chi^{''} \tag{5.10}$$

where real and complex components are given by

$$\begin{aligned} \chi^{'}(\omega) &= \frac{\mu_{ij}^2(\omega_{ij}-\omega)T_{ij}}{2\varepsilon_0\hbar}\Delta N_{ij}L_{ij}(\omega), \\ \chi^{''}(\omega) &= \frac{\mu_{ij}^2}{2\varepsilon_0\hbar}\Delta N_{ij}L_{ij}(\omega), \end{aligned} \tag{5.11}$$

where μ_{ij} is the dipole matrix element, ω_{ij} is the angular frequency, T_{ij} is the dephasing time, ΔN_{ij} is the population inversion and $L(\omega)$ is the frequency dependent Lorentzian line shape, all for the transition $i \rightarrow j$. The Lorentzian is defined as

$$L_{ij}(\omega) = \frac{\gamma_{ij}}{\pi\left\{(\hbar\omega - |E_{ij}|)^2 + \gamma_{ij}^2\right\}}, \tag{5.12}$$

where γ is the linewidth, defined as $\gamma_{ij} = \frac{\hbar}{T_{ij}}$, and E is the transition energy, both for the transition $i \rightarrow j$.

While the lasing transition will have the biggest contribution to the complex refractive index at lasing frequency, in order to find the total frequency-dependent refractive index (5.11) must be summed over all transitions, i.e.

$$\chi_{\text{total}} = \sum_{i,j}\chi_{ij}, \qquad \text{for } j > i.$$

The total permittivity is the sum of the permittivity of the host material plus the calculated electoral contribution, i.e.

$$\begin{aligned} \epsilon_{\text{Total}} &= \epsilon_{\text{Host}} + \epsilon_{\text{Elec.}}, \\ &= \epsilon_{\text{Host}}(1 + \frac{\chi}{\epsilon_{\text{Host}}}), \end{aligned} \tag{5.13}$$

and the complex refractive index of the form $\mu + \mathrm{i}\|$ is the square root of the permittivity, such that

$$\begin{aligned} \mu + \mathrm{i}\| &= \sqrt{\epsilon_{\text{Host}}}\sqrt{(1 + \frac{\chi}{\epsilon_{\text{Host}}})}, \\ &\approx \sqrt{\epsilon_{\text{Host}}}(1 + \frac{\chi}{\epsilon_{\text{Host}}}), \end{aligned} \tag{5.14}$$

Figure 5.13 shows the frequency dependent gain , and 5.14 shows the frequency dependent refractive index, calculated for the three-well active region QCL by Luo *et al.* [38]. The gain peak is at the correct frequency of 3.2 THz with a reasonable amplitude with the other features in the plot coming from the other transitions away from the lasing frequency. The

main inflection of the refractive index (at 3.2 THz) also appears correct, with the refractive index increasing below the lasing frequency and decreasing above. The features from the other transitions within the QCL are much more prominent in the refractive index, particularly the inflection ∼1 THz which corresponds to the transition between the lower laser level and the extraction state. This is due to the large dipole matrix element between the two states (the two states anti-cross at around the operating voltage of 12.5kV/cm).

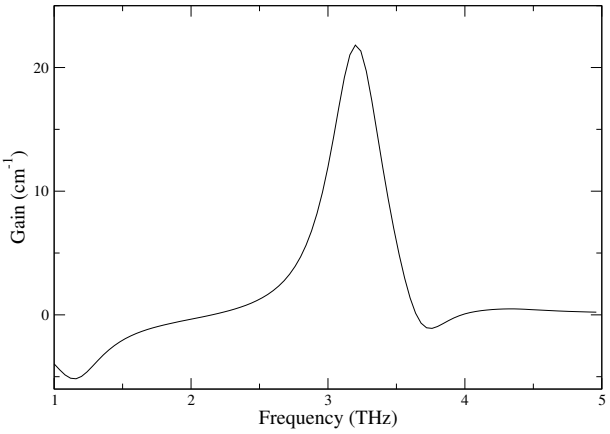

Figure 5.13: Frequency dependent gain of the three-well active region structure by Luo *et al.* [38] with an applied bias of 12.5 kVcm^{-1} (i.e. the design bias).

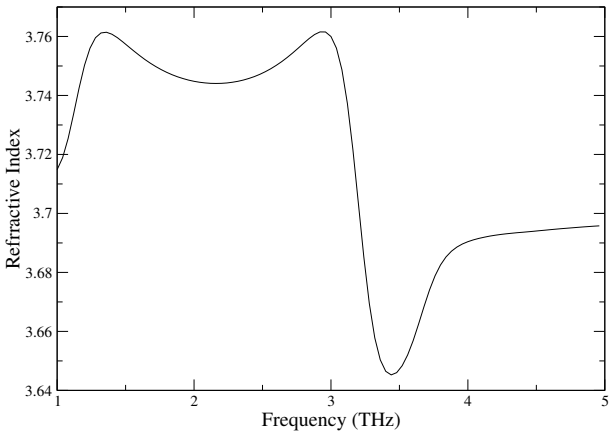

Figure 5.14: Frequency dependent total refractive index (i.e. both host and electrical contributions) for the three-well active region structure by Luo *et al.* [38] with an applied bias of 12.5 kVcm^{-1} (i.e. the design bias).

5.2.4 Change in carrier concentration

As a change in carrier concentration will effect the magnitude charge distribution with the active region of the QCL, and since the solution of the electronic structure is self-consistent and dependent upon the carrier concentration, a change in the carrier concentration must be considered within the self-consistent loop of the rate equation model. Fortunately this is not difficult, owing to the fact that the donor atoms do not move and their concentration does not change, and that the number of free carriers is simply distributed amongst the states within the QCL according to the ratios found by solving the rate equations. Therefore a change in carrier concentration can be implemented by changing the number of free carriers available to distribute between states, although importantly not changing the number, and therefore the charge from the dopant atoms.

To simulate the effect of a depletion of carriers caused by the SAW induced potential interacting with the QCL active region, a range of different carrier concentration where given to the rate equation model to determine the change in gain and refractive index with varying carrier concentration. Figure 5.15 shows the change in gain and 5.16 shows the change in refractive index for a range of carrier concentrations. These do not scale linearly with the carrier concentration due to the self-consistent loop as well as, in particular, the inclusion of electron-electron scattering.

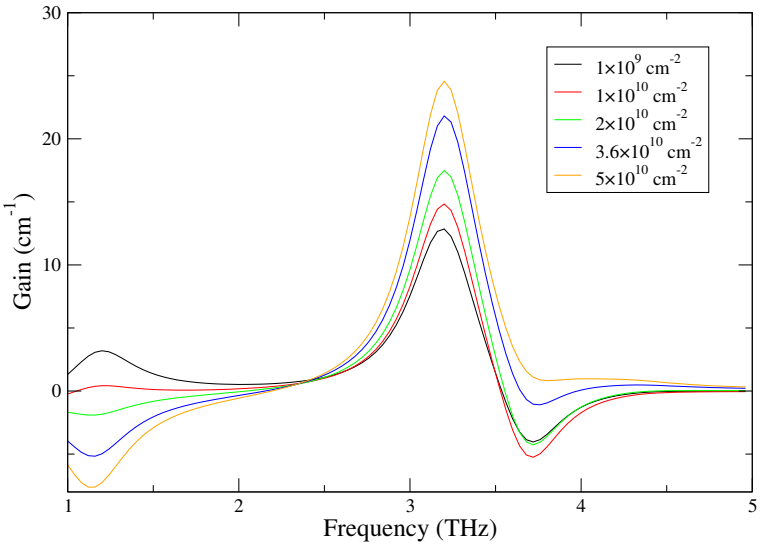

Figure 5.15: Gain of the three-well active region structure by Luo *et al.* [38], with the applied bias set at the design bias of 12.5 kVcm^{-1}, for a range of carrier concentrations, with the *normal* carrier concentration being 3.6×10^{10} cm$^{-}2$.

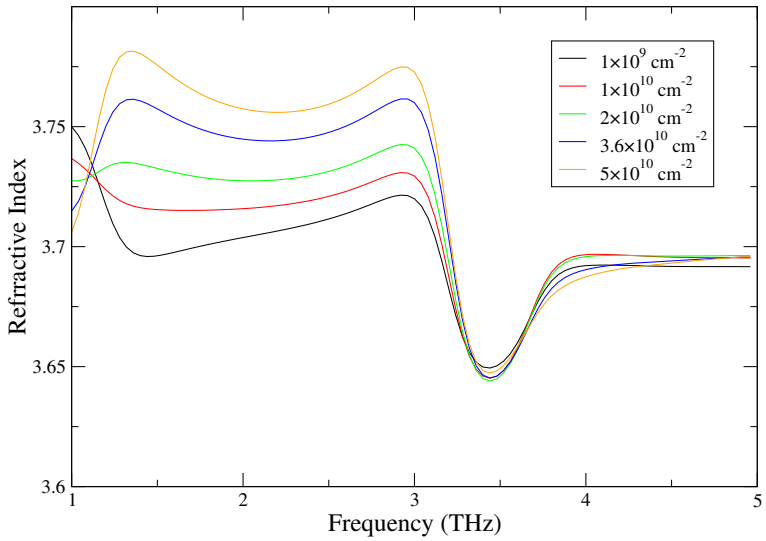

Figure 5.16: Refractive index of the three-well active region structure by Luo *et al.* [38], with the applied bias set at the design bias of $12.5\,\mathrm{kVcm}^{-1}$, for a range of carrier concentrations, with the *normal* carrier concentration being $3.6{\times}10^{10}\,\mathrm{cm}^{-}2$.

5.3 Design recommendation for a SAW modulated QCL

As shown in section 5.1.2, the biggest problem associated with the SAW modulated QCL design presented by Salih *et al.* [36] is the discontinuity in the surface of the structure between the substrate and the QCL ridges. As discussed, this discontinuity causes an unknown but potentially significant amount of scattering of the incident SAW energy into a bulk propagation mode which will not enter the QCL ridge. This means that not only can the modulation of the carrier concentration in the QCL AR not be determined by the model presented in this thesis, but the magnitude of that modulation is potentially very small. Furthermore, the sloped facets on the QCL ridge reduce its effectiveness as a resonant cavity which degrades the performance of the QCL as discussed in [36].

As a result of these problems, a recommended design is presented which removes the need for the SAW to propagate over any discontinuities in the surface structure. Not only does this remove the problems involved with simulating ridge structures, thereby allowing the revised device design to be modelled using the simulation tools presented in this thesis, but also removes the biggest unknown factor in the design from [36]; that is, how much of the SAW energy is scattered by the discontinuity of the surface of the device.

The presented deign comes in two *flavours* — the need for which is discussed presently, each of which use the same principle of placing an IDT on top of the QCL cavity and generating the SAW directly within the QCL AR. This removes the need for the SAW to move from the substrate into the QCL cavity and the requirement for the QCL facets to be sloped which reduces the effectiveness of the laser cavity. (Furthermore, the fabrication of the device will be simpler as only one additional step beyond the QCL fabrication is required, i.e. patterning an IDT on the top contact, as opposed to fabricating the QCL and mounting on a SAW device as discussed in [36].)

The first design is shown in figure 5.17. Within this design, an IDT structure is patterned along the entire length of the QCL ridge in place of the top contact metallisation. Since the QCL is driven by a relatively DC voltage compared to the IDT voltage which is an RF signal, these two signals may be added such the applied signal may be used to both electrically pump the QCL and generate a SAW (with the RF signal on one side of the IDT out of phase by π from the opposite side). The main drawback of this method is that, because of the n-doped top-contact layer on the QCL which is partially conductive, some of the RF voltage applied to the IDT will leak across to the opposite side of the IDT, reducing the effectiveness of the IDT at producing a SAW. While the magnitude of the RF voltage can be increased to counteract this effect, the limiting factor is how much oscillation in its driving current the QCL can withstand whilst still maintaining normal operation. Since the effect of a conductive substrate reducing the effectiveness of the IDTs to generate a SAW is not considered within the current model[2], the magnitude of the required RF voltage cannot be calculated. Instead, a second design is presented which may be used to sidestep this problem if it becomes a significant issue.

Figure 5.18 shows structure of the second SAW modulated QCL suggested design. Within this design, a section of the QCL cavity is electrically isolated by depositing an insulating layer in place of the n-doped top contact layer. The rest of the laser is fabricated with the *normal* top-contact metallisation n-doped top contact layer. This has the effect of only part of the laser cavity is electrically pumped (the part underneath the normal top-contact metallisation) and therefore contributing to gain whilst the rest of the cavity (the section underneath the IDT) does not act as a gain medium but only as a waveguide. Whilst effectively *switching off* part of the QCL cavity will reduce the overall gain, and therefore the output power, of the QCL, doing so does allow the driving voltages for the QCL and IDT to be completely separated. However, as shown in chapter 4.3, the magnitude of the generated is dependent

[2]This is set as further work in chapter 6.1.

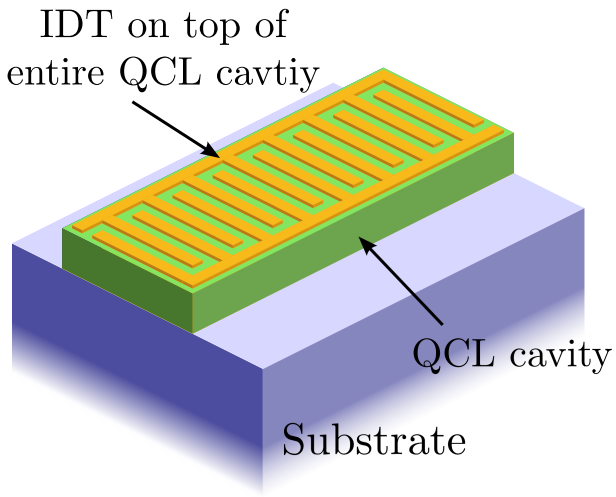

Figure 5.17: Illustration of the first design recommendation for a SAW modulated QCL. The IDT sits on top of the QCL cavity and runs for the entire length of the QCL cavity ensuring that a SAW is generated evenly throughout the QCL cavity.

upon the number of finger pairs, and therefore the overall length of the generating IDT. There is, therefore, a trade-off within this design between how much SAW power is needed (i.e. how long the IDT must be) and how much the gain of the QCL may be reduced by (i.e. how much of the QCL may be switched off). Furthermore, within this design the loss in SAW power as it travels along the QCL ridge will be an important effect as the SAW will have to travel a significant distance — of the order of several millimetres.

While the presented model cannot be used to simulate the effect of the conductive top contact layer on SAW generation in the first design, or the loss of SAW power as it travels along the QCL ridge in the second design (as the IDT model assumes the two sides of the IDT are electrically isolated and the FDTD acoustic wave model is losses), it can be used to find an approximation to the modulation in carrier concentration of the QCL without these effects. Therefore the SAW model was used in conjunction with the rate-equation model to simulate the Luo *et al.* QCL active region, as used in the Salih *et al.* SAW-modulated QCL design, with various carrier concentrations determined by the level of modulation in the carrier concentration from the SAW (as in section 5.2.4). The wavelength of the SAW was set to 12.5 μm corresponding to half the wavelength of the simulated emission frequency ($\sim$3.2 THz) within the gain medium (such that the modulating frequency satisfies the Bragg condtion [62]) and the voltage on each side of the IDT was set to $\pm 1\,\mu$V. The conductive layers

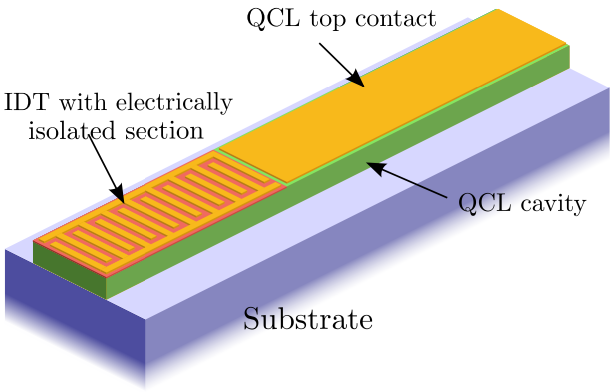

Figure 5.18: Illustration of the second design recommendation for a SAW modulated QCL. Part of the QCL cavity is has an IDT patterned on top whilst the rest is has a metallised top-contact. The area underneath the IDT electrically isolated from the QCL active region to stop the potential applied to the IDT interfering with the electrical pumping of the QCL. The SAW is therefore generated within part of the QCL which is effectively *switched-off* before propagating into, and modulating, the operational part.

representing the QCL active region within the SAW model were placed at depths between 0.5 μm and 15.5 μm below the surface of the QCL ridge, giving a top contact layer thickness of 500 nm and a total AR thickness of 15 μm. The unperturbed carrier concentration within these layers was set to that of the Luo *et al.* QCL, i.e. $8.1 \times 10^{14}\,\mathrm{cm}^{-1}$. This gave results for the modulation in carrier concentration very similar to those in section 5.2.4, with a maximum carrier concentration of $2.5 \times 10^{15}\,\mathrm{cm}^{-1}$ and the minimum being fully depleted areas. This range of carrier concentrations was used within the rate-equation model to calculate the complex refractive index for the QCL AR for various carrier concentrations, giving identical results to those shown in section 5.2.3.

In order to give a quantitative discussion as to whether this level of modulation will lead to DFB within the QCL, these results must be analysed using coupled-wave theory [62]. While a full analysis is beyond the scope of this thesis (and is set as further work in section 6.1), the DFB coupling constant may be easily calculated and used as an indicator as to whether DFB is likely. From [62], a periodic modulation in the refractive index and gain of laser gain medium may be described as

$$n(x) = n_0 + n_1 \cos 2\beta_0 x \tag{5.15}$$

$$g(x) = g_0 + g_1 \cos 2\beta_0 x, \tag{5.16}$$

respectively, where x is the long-axis of the QCL (and the propagation direction of the SAW), $n(x)$ is spatially varying refractive index with n_0 and n_1 being the unperturbed value and magnitude of the modulation of refractive index respectively, $g(x)$ is the spatially varying gain with g_0 and g_1 being the unperturbed value and magnitude of the modulation of gain respectively, and β_0 is given as

$$\beta_0 = \frac{2\pi n_0}{\lambda_0}, \tag{5.17}$$

where λ_0 is the periodicity of the modulation to the gain and refractive index, i.e. the wavelength of the SAW, which in this case is 12.5 μm. The DFB coupling constant, κ, is then given as

$$\kappa = \frac{\pi n_1}{\lambda_0} + \frac{1}{2} j g_1. \tag{5.18}$$

Figure 5.19 shows the frequency dependent coupling constant, $|\kappa|$, for the simulated device. At the simulated emission frequency, 3.2 THz, the magnitude of the coupling constant is 9.0 cm^{-1}. Comparing this value to others in the literature for DFB QCLs such as [26, 67], which both have values for the $|\kappa|$ of 5 cm^{-1}, implies that the strength of the modulation from the SAW is more than strong enough to induce DFB within the QCL.

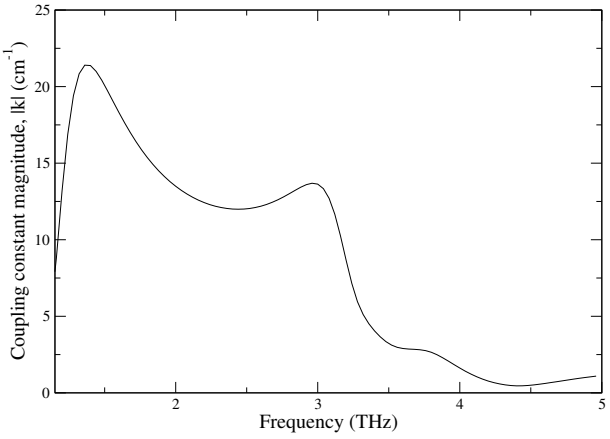

Figure 5.19: Simulated frequency-dependent coupling constant for the Luo *et al.* QCL AR [38] with an applied bias of 12.5 kVcm^{-1} (i.e. the design bias) modulated by a SAW with a wavelength of 12.5 μm (i.e. the frequency which satisfies the Bragg condition [62] for the QCL emission wavelength, 3.2 THz).

Chapter 6

Conclusion

A set of simulation tools for modelling the modulating effect of a surface acoustic wave (SAW) on the carrier concentration within a quantum cascade laser (QCL) active region have been developed. These tools have been used to find the expected modulation in complex refractive index within the QCL AR caused by a SAW, from which the distributed feedback (DFB) coupling constant has been calculated. While the presented model has not been able to model SAW modulated QCL device design that was proposed and fabricated by Salih *et al.* [36] due to the shape of the surface structure causing instabilities within the simulation, it was concluded that these instabilities suggest that a significant portion of the SAW energy was scattered by the discontinuity where the substrate surface meets the QCL ridge and therefore much of the SAW would not contribute to modulating the QCL AR carrier concentration. An alternative design has therefore been suggested which not only can be modeled by the presented tools, but more importantly removes any uncertainty of how much of the SAW will modulate the carrier concentration by generating the SAW directly within the QCL ridge. For this alternate design, the magnitude of the coupling constant, $|\kappa|$, at the emission wavelength was found to be $\sim 9\,\mathrm{cm}^{-1}$, indicating that DFB modulation of QCLs using SAWs is indeed feasible. Further work which would help to validate this claim is presented at the end of this chapter.

The self-consistent Schrödinger-Poisson solution provides an accurate method for determining the electrical structure of semiconductor heterostructures including the effects of charge. The accuracy and reliability of this method were improved by the derivation of a novel, direct solution to the nonparabolic Schrödinger equation in chapter 2. This allows material systems with high conduction band offsets (of the order of 1 eV) to be modelled without the problem of missing pairs of states which are energetically very close which is

often suffered by iterative approaches. This is particularly appealing for the design of QCLs where, due to the very narrow barrier widths, states are often very close in energy and the design work may be automated (through a genetic algorithm or similar) requiring the solution to be robust, i.e. not skip states. The diagonalisation method for solving Poisson's equation was also investigated, and a new diagonalisation algorithm was derived for the problem with mixed boundary conditions. Furthermore, it is thought that utilising the same optimisation generating algorithm (appendix B) that a similar diagonalisation algorithms could be produced for other sparse banded matrices, such as the matrix produced from the discretisation of the two-dimensional Poisson equation.

The derivation of a model of SAW propagation requires an in depth understanding of acoustic wave propagation in crystalline materials. Therefore the acoustic wave equations of motion for propagation within piezoelectric crystals were derived from first principles in chapter 3. Since any simulation acoustic wave propagation will inherently be an open-domain problem, it will suffer from the problem of energy within the system reflecting of simulation domain boundaries and reflecting back into the region of interest; a problem that was previously unsolved within piezoelectric crystals. This was solved by the application of the concept of PMLs, predominantly used in electromagnetic simulations, to the acoustic wave equations of motion.

In order to simulate SAWs, a surface boundary condition has to be added to the acoustic wave simulation. This was derived in chapter 4. Furthermore, as SAWs are excited using IDT structures experimentally, the time-dependent potential around an IDT structure was modelled (by introducing the ability to fix the potential within the two-dimensional Poisson's equation), and this potential was used to excite the SAW simulation. The simulated SAWs compared well to the analytical solution for a SAW, and had an almost identical value of propagation velocity (and therefore wavelength) as measured experimentally. Additionally, several SAW devices were fabricated with various numbers of finger pairs in the two IDTs on the device (fabrication and processing done by Dr. M. Salih), in order to exploit the change in FWHM of the S21 response of the device with varying numbers of finger pairs. These devices were simulated and the responses compared, showing an excellent agreement in the variation of FWHM with varying numbers of fingers. While not utilised within this project, this novel method for simulating real-world sized IDT devices may act as a design tool for determining the response of arbitrary IDT structures in two dimensions.

To examine the effect of SAW propagation through a QCL ridge the FDTD acoustic-wave simulation had to be extended to allow modelling a QCL AR. Most notably, this included

the addition of a conductive layer within the substrate which mimics the effect of the free carries within the QCL AR. In order to simulate the design proposed by Salih *et al.* which included a QCL ridge bound to the top of the SAW substrate, discontinuities in the surface of the acoustic wave simulation were investigate, however a suitable boundary condition for the corner points of this ridge was not found that did not introduce instabilities into the simulation. Since it was already assumed that the discontinuity in the surface would cause some scattering of the SAW (hence why this design employs several features which attempt to minimise this scattering), it was concluded that these instabilities imply that a significant amount of scattering would occur and that they were only unstable within the simulation because the structure being simulated was not physically realistic (i.e. contained exact right-angled corners). Furthermore, the self-consistent Schröindger-Poisson model was extended to scattering rate model allowing the populations of all states within the QCL AR to found and a method for calculating the complex refractive index which utilises these populations was presented. By varying the carrier concentration used within the rate equation over the range predicted within the AR by the SAW-QCL simulation, the variation in the frequency dependent complex refractive index can be calculated for a SAW modulated QCL using the presented model. Finally, a recommended design was presented that removes any issues of acoustic-wave scattering stopping the SAW modulating the QCL AR by generating the SAW directly within the QCL ridge itself. The presented model was used to calculate the change in carrier concentration within the QCL AR for the proposed design, from which the variation in complex refractive index was then calculated. The DFB coupling constant was then calculated for this design as $\sim 15\,\mathrm{cm}^{-1}$, implying that the SAW modulation would be strong enough to induce DFB within the QCL.

6.1 Further work

The presented analysis of the modulation in QCL AR carrier concentration by a SAW has indicated that this is a feasible method of achieving tunable, DFB modulation of a QCL. The model, however, is by no means complete and there are several avenues of investigation which could not only help to validate this conclusion, but also expand upon the results presented within this thesis.

Firstly, the current model has not been exhaustively to explore how different QCL ARs perform over a range of SAW frequencies. The analysis presented in section 5.3 was only used to validate if the SAW modulation was strong enough at the resonant frequency of the IDT for

the specific QCL AR discussed. This line of investigation could be deepened substantially by examining how the DFB coupling constant changes as the frequency of the oscillating voltage applied to the IDT is moved away from the IDT resonant frequency. Comments could then be made about the expected tuning range of this structure. Furthermore, this analysis could be done with QCL AR designs, each of which will have different doping densities and therefore different *unperturbed* carrier concentrations resulting in different levels of SAW modulation. Different QCL ARs could they be assessed as to which is most favourable to SAW modulation, by which has the highest DFB coupling constants.

Next, the assumptions and approximations within the model itself could be improved. The most prominent assumption for the recommended design is that the n-doped top contact layer of the QCL will not have a significant impact on the IDTs ability to generate a SAW. In order to rectify this assumption the IDT model will need to be expanded to deal with a conductive substrate. There is also the issue of loss within the SAW simulation itself. As stated for second *flavour* of the recommended design, propagation losses in SAW energy may be significant as the SAW may be required to travel a significant distance along the QCL ridge. Loss may be added to FDTD acoustic wave model by adding a loss term to the material constants, as discussed in [12].

Finally, by using calculated DFB coupling constant as an input into coupled wave theory [62, 68], the modulation of the spectral output from the QCL can be determined given a particular strength and depth of modulation of the carrier concentration through the QCL AR. This would allow the strength of the modulation required to achieve some frequency tuning range to be determined, and allow the methods presented within this thesis to be used as a design tool for improving SAW modulation of QCLs.

Because of the generality of the SAW model, there are several other experimental validations which could done with the model. For example, if the model was used as an IDT design tool then any new IDT designs could be fabricated and their measured responses compared with the predicted ones. Once the reliable simulation of ridge structures becomes a reality, experimental validation could also be used to confirm that the simulated SAW propagation is correct by fabricating devices where the receiving IDT is on top of a ridge of varying heights. The presence of grooves etched into the surface could also be investigated and used for model validation, as well as periodic surface structures which could induce a resonant effect with the SAW similar to a Bragg grating.

Due to there similarity to SAW-QCL devices, acoustic charge transport (ACT) devices could be modelled using the outlined method by extending the model of buried conductive

layers. ACT devices use a propagating SAW to transport charge within a 2DEG buried below the surface of the substrate. Often the charge is transported through some form of gate, consisting of a series of metal electrodes patterned on the surface which, when a voltage is applied, will deplete the 2DEG underneath of carriers [64]. With the right geometry, the gate can be made to pinch off the 2DEG such that within the gate area the electrons are confined to a one dimensional channel [69]. These single-electron transport devices have a broad range of applications including the study of fundamental quantum mechanics [70], quantum metrology [71] and quantum computing [72]. While the ability to model time dependent transport of electrons would need to be added, the SAW model has all the other functionality it would need and ability to simulate the confinement of electrons in a one-dimensional channel is already provided by the Schrödinger-Poisson solution. Such a solution would be in contrast to other theoretical investigations into ACT which have utilised an analytical solution to the SAW equations of motion to define the potential of the SAW which contains the transported charge packet [18, 19, 20]. Furthermore, this model could also be used to examine acoustic charge transport in graphene placed on the surface of a SAW device. These devices are, whilst relatively new in their realisation, have the potential to be used to investigate fundamental quantum effects due to the interesting electronic structure of the mono-layer of graphene [73].

Appendix A

Finite difference relations

The first derivative of some function is the gradient over some small distance δx,

$$\left.\frac{\partial f(x)}{\partial x}\right|_i \approx \frac{f_{i+1} - f_{i-1}}{\delta x}. \tag{A.1}$$

For the second derivative,

$$\begin{aligned}
\frac{\partial^2 f(x)}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial f(x)}{\partial x}\right), \\
\left.\frac{\partial f(x)}{\partial x}\right|_i &\approx \frac{\left.\frac{\partial f(x)}{\partial x}\right|_{i+1} - \left.\frac{\partial f(x)}{\partial x}\right|_{i-1}}{\delta x}, \\
&\approx \frac{\frac{f_{i+2} - f_i}{\delta x} - \frac{f_i - f_{i-2}}{\delta x}}{\delta x}, \\
&\approx \frac{f_{i+2} - 2f_i + f_{i-2}}{\delta x}.
\end{aligned}$$

Removing a factor of 2 from δx [41], gives

$$\left.\frac{\partial f(x)}{\partial x}\right|_i \approx \frac{f_{i+1} - 2f_i + f_{i-1}}{\delta x}. \tag{A.2}$$

Appendix B

Optimisation algorithm for matrix digitalisation

The optimisation algorithm for producing consists of performing Gaussian elimination on the matrix to reduce it to an upper triangle, whilst documenting the arithmetic performed on each non-zero element of the matrix. This documenting is done using the String class within the C++ standard library since it facilitates an easy *search and replace* method to replace strings of characters.

The optimisation begins by defining the form of the matrix to produce an optimisation algorithm for, i.e. defining the positions of the non-zero elements. This is done for a very small matrix to reduce the complexity of the output form the optimisation algorithm, usually 6×6, although depending on the positions of the bands it may have to be larger. As each *move* of the digitalisation is performed, an equation (stored as a string) is created from the elements involved in the digitalisation (this will a ratio of two elements multiplied by the element which is being changed and stored in the position of the changed element. If the elements involved in that move or the updated element is zero then no equation is stored. On subsequent *moves*, new equations are made up from the equations stored from previous moves such that at the end of the digitalisation each element on the main diagonal contains an equation describing exactly how that element was made. Even with a small matrix these equations become very large, containing a large number of terms, making is almost impossible to find patterns within them and simplify these equations *by eye*.

Fortunately, the search and replace feature of the String class can be used for this task. Starting with the first element of the main diagonal (which will be unchanged), the equation for the current element is searched for, and replaced within the equation for the next ele-

ment for a symbol which defines the previous equation. This reduces the complexity of the equations enough such that patterns may be easily spotted by eye, and a custom algorithm for diagonalising this form of matrix may be written from these equations. This custom algorithm effectively removes the redundant steps from the Gaussian elimination which result from the zero elements of banded matrices.

Appendix C

Material constants

C.1 Material constants for acoustic wave simulation

C.1.1 Gallium arsenide, [1,0,0] propagation direction

$$\mathsf{C} = \begin{pmatrix} 11.9 \times 10^{10} & 5.34 \times 10^{10} & 5.34 \times 10^{10} & 0 & 0 & 0 \\ 5.34 \times 10^{10} & 11.9 \times 10^{10} & 5.34 \times 10^{10} & 0 & 0 & 0 \\ 5.34 \times 10^{10} & 5.34 \times 10^{10} & 11.9 \times 10^{10} & 0 & 0 & 0 \\ 0 & 0 & 0 & 5.96 \times 10^{10} & 0 & 0 \\ 0 & 0 & 0 & 0 & 5.96 \times 10^{10} & 0 \\ 0 & 0 & 0 & 0 & 0 & 5.96 \times 10^{10} \end{pmatrix},$$

$$\mathsf{e} = \begin{pmatrix} 0 & 0 & 0 & -0.16 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.16 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.16 \end{pmatrix},$$

$\varepsilon = 97.4 \times 10^{-12}$,

$\rho = 5317.$

C.1.2 Gallium arsenide, [1,1,0] propagation direction

$$\mathsf{C} = \begin{pmatrix} 1.458 \times 10^{11} & 2.66 \times 10^{10} & 5.34 \times 10^{10} & 0 & 0 & 0 \\ 2.66 \times 10^{10} & 1.458 \times 10^{11} & 5.34 \times 10^{10} & 0 & 0 & 0 \\ 5.34 \times 10^{10} & 5.34 \times 10^{10} & 11.9 \times 10^{10} & 0 & 0 & 0 \\ 0 & 0 & 0 & 5.96 \times 10^{10} & 0 & 0 \\ 0 & 0 & 0 & 0 & 5.96 \times 10^{10} & 0 \\ 0 & 0 & 0 & 0 & 0 & 3.28 \times 10^{10} \end{pmatrix},$$

$$\mathsf{e} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0.16 & 0 \\ 0 & 0 & 0 & -0.16 & 0 & 0 \\ 0.08 & -0.08 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$\varepsilon = 97.4 \times 10^{-12}$,

$\rho = 5317$.

C.1.3 Lithium niobate, [1,0,0] propagation direction

$$\mathsf{C} = \begin{pmatrix} 2.03 \times 10^{11} & 0.573 \times 10^{11} & 0.752 \times 10^{11} & 0.085 \times 10^{11} & 0 & 0 \\ 0.573 \times 10^{11} & 2.03 \times 10^{11} & 0.752 \times 10^{11} & -0.085 \times 10^{11} & 0 & 0 \\ 0.573 \times 10^{11} & 0.752 \times 10^{11} & 2.424 \times 10^{11} & 0 & 0 & 0 \\ 0.085 \times 10^{11} & -0.085 \times 10^{11} & 0 & 0.595 \times 10^{11} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.595 \times 10^{11} & 0.085 \times 10^{11} \\ 0 & 0 & 0 & 0 & 0.085 \times 10^{11} & 0.7285 \times 10^{11} \end{pmatrix},$$

$$\mathsf{e} = \begin{pmatrix} 0 & 0 & 0 & 0 & 3.76 & -2.43 \\ -2.43 & 2.43 & 0 & 3.76 & 0 & 0 \\ 0.23 & 0.23 & 1.33 & 0 & 0 & 0 \end{pmatrix},$$

$\varepsilon = 0.377 \times 10^{-9}$,

$\rho = 4640$.

C.1.4 Lithium niobate, Y-cut surface, 128° rotated. ([-0.78, 0, 1] propagation direction)

$$\mathsf{C} = \begin{pmatrix} 2.03 \times 10^{11} & 7.52 \times 10^{10} & 5.73 \times 10^{10} & -7.77 \times 10^{9} & 0 & -3.46 \times 10^{9} \\ 6.84 \times 10^{10} & 2.42 \times 10^{11} & 6.41 \times 10^{10} & 0 & -8.68 \times 10^{9} & 0 \\ 5.73 \times 10^{10} & 7.52 \times 10^{10} & 2.03 \times 10^{11} & 7.77 \times 10^{9} & 0 & 3.46 \times 10^{9} \\ -7.77 \times 10^{9} & 0 & 7.77 \times 10^{9} & 5.95 \times 10^{10} & 3.46 \times 10^{9} & 0 \\ 0 & 0 & 0 & 3.46 \times 10^{9} & 7.29 \times 10^{10} & -7.77 \times 10^{9} \\ -3.46 \times 10^{9} & 0 & 3.46 \times 10^{9} & 0 & -7.77 \times 10^{9} & 5.95 \times 10^{10} \end{pmatrix},$$

$$\mathsf{e} = \begin{pmatrix} 0.26 & 0 & -0.26 & 0 & 4.078 & 3.76 \\ 0.23 & 1.33 & 0.23 & 0 & 0 & 0 \\ 1.29 & 0 & -1.29 & 3.76 & -2.44 & 0 \end{pmatrix},$$

$\varepsilon = 0.377 \times 10^{-9}$,

$\rho = 4640$.

C.1.5 Bismuth germinate, [1,0,0] propagation direction

$$\mathsf{C} = \begin{pmatrix} 11.58 \times 10^{10} & 2.7 \times 10^{10} & 2.7 \times 10^{10} & 0 & 0 & 0 \\ 2.7 \times 10^{10} & 11.58 \times 10^{10} & 2.7 \times 10^{10} & 0 & 0 & 0 \\ 2.7 \times 10^{10} & 2.7 \times 10^{10} & 11.58 \times 10^{10} & 0 & 0 & 0 \\ 0 & 0 & 0 & 4.36 \times 10^{10} & 0 & 0 \\ 0 & 0 & 0 & 0 & 4.36 \times 10^{10} & 0 \\ 0 & 0 & 0 & 0 & 0 & 4.36 \times 10^{10} \end{pmatrix},$$

$$\mathsf{e} = \begin{pmatrix} 0 & 0 & 0 & -0.0367 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.0367 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.0367 \end{pmatrix},$$

$\varepsilon = 141.67 \times 10^{-12}$,

$\rho = 7095$.

C.1.6 Bismuth germinate, [1,0,0] propagation direction

$$\mathsf{C} = \begin{pmatrix} 1.15 \times 10^{11} & 2.78 \times 10^{10} & 2.7 \times 10^{10} & 0 & 0 & 0 \\ 2.78 \times 10^{10} & 1.15 \times 10^{11} & 2.7 \times 10^{10} & 0 & 0 & 0 \\ 2.7 \times 10^{10} & 2.7 \times 10^{10} & 11.58 \times 10^{10} & 0 & 0 & 0 \\ 0 & 0 & 0 & 4.36 \times 10^{10} & 0 & 0 \\ 0 & 0 & 0 & 0 & 4.36 \times 10^{10} & 0 \\ 0 & 0 & 0 & 0 & 0 & 4.44 \times 10^{10} \end{pmatrix},$$

$$\mathsf{e} = \begin{pmatrix} 0 & 0 & 0 & 0 & -0.0367 & 0 \\ 0 & 0 & 0 & -0.0367 & 0 & 0 \\ -0.01835 & -0.01835 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$\varepsilon = 141.67 \times 10^{-12}$,

$\rho = 7095$.

Appendix D

Einstein summation convention

The Einstein summation convention allows sets of equations which reuse the same terms, only with different subscripts, to be written in a more succinct manner than writing them out in full. It is used often in equations describing physical effects in multiple dimensions, particular those containing higher rank tensors such as Hooke's law, as it provides an accurate way to portray all the information contained within the equation using a little notation as possible.

Consider the relation

$$\mathsf{A} = \mathsf{BC}. \tag{D.1}$$

While the form of this relation is clear, it is not fully described unless the form of each of the three elements is defined. Suppose that A and C are both rank 1 tensors (i.e. a vector), B is a rank 2 tensor (i.e. a matrix), and the equation is applied in 3-dimensional space, then it may be written out more fully as

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}. \tag{D.2}$$

Equation (D.2) actually represents three separate equations, which are

$$\begin{aligned} a_1 &= b_{11}c_1 + b_{12}c_2 + b_{13}c_3, \\ a_2 &= b_{21}c_1 + b_{22}c_2 + b_{23}c_3, \\ a_3 &= b_{31}c_1 + b_{32}c_2 + b_{33}c_3. \end{aligned} \tag{D.3}$$

Notice that each of these equations is identical in its form, except for changes within each subscript and furthermore that each term is identical, except for a change in subscript. Because of the similarity between these equations and terms within the equations, the information they represent may be written more succinctly by *generalising* the subscripts they use. Firstly,

the three equations may be written as one equation by generalising the first subscript of B, that is

$$a_i = b_{i1}c_1 + b_{i2}c_2 + b_{i3}c_3 \qquad \text{for } i = 1, 2, 3. \tag{D.4}$$

Secondly, the second subscript on B may be generalised and, using a summation, the three separate terms gathered into one, such that

$$a_i = \sum_{j=1}^{3} b_{ij}c_j \qquad \text{for } i = 1, 2, 3. \tag{D.5}$$

Equation (D.5) has all of the information from (D.2) only written in a much more succinct fashion.

The Einstein summation convention is method of further simplifying the notation by removing the need for the summation since it is implied by convention. It states that whenever there is a repeated subscript within a term of an equation, that this implies a summation of that term over that subscript, and whenever there is single, not repeated subscript within a term, this implies separate equations with the same terms, different values for that subscript in each equation. Using this convention, (D.5) may be written even more succinctly as

$$a_i = b_{ij}c_j \qquad \text{for } i, j = 1, 2, 3. \tag{D.6}$$

Often, although it is not strictly part of the convention, the definition of the subscripts is dropped and the equation be be written as

$$a_i = b_{ij}c_j. \tag{D.7}$$

Interestingly, (D.7) is actually more general the (D.2) since its subscripts are not bound to any particular limits. Furthermore, while (D.7) has all the generality of (D.1) (the implication being that $i, j = 1, \ldots, n$ where n is number of dimension within the system), (D.7) actually contains more information than (D.1) since the form of its variables are all definitely defined. Therefore in writing (D.2) in a more succinct manner, the Einstein summation convention has actually produced an equation which contains more information than the original, general relation.

The Einstein summation convention is used often in equations describing physical effects in multiple dimensions, particular those containing higher rank tensors such as Hooke's law, as it provides an accurate way to portray a lot of information in a very succinct manner.

Bibliography

[1] Jérôme Faist, Federico Capasso, Deborah L. Sivco, Carlo Sirtori, Albert L. Hutchinson, and Alfred Y. Cho. Quantum cascade laser. *Science*, 264:553–556, April 1994.

[2] Noriaki Horiuchi. Terahertz technology: Endless applications. *Nature Photonics*, 4(3):140–140, 2010.

[3] Masayoshi Tonouchi. Cutting-edge terahertz technology. *Nature photonics*, 1(2):97–105, 2007.

[4] E Pickwell and VP Wallace. Biomedical applications of terahertz technology. *Journal of Physics D: Applied Physics*, 39(17):R301, 2006.

[5] Peter H Siegel. Terahertz technology. *Microwave Theory and Techniques, IEEE Transactions on*, 50(3):910–928, 2002.

[6] Michael C Kemp, PF Taday, Bryan E Cole, JA Cluff, Anthony J Fitzgerald, and William R Tribe. Security applications of terahertz technology. In *AeroSense 2003*, pages 44–52. International Society for Optics and Photonics, 2003.

[7] J Faist, F Capasso, C Sirtori, DL Sivco, JN Baillargeon, AL Hutchinson, SNG Chu, and AY Cho. High power mid-infrared (lambda greater than or similar to-5 mu m) quantum cascade lasers operating above room temperature. *APPLIED PHYSICS LETTERS*, 68(26):3680–3682, JUN 24 1996.

[8] JS Yu, A Evans, J David, L Doris, S Slivken, and M Razeghi. High-power continuous-wave operation of quantum-cascade lasers up to 60 degrees C. *IEEE PHOTONICS TECHNOLOGY LETTERS*, 16(3):747–749, MAR 2004.

[9] S Fathololoumi, E Dupont, CWI Chan, ZR Wasilewski, SR Laframboise, D Ban, A Mátyás, C Jirauschek, Q Hu, and HC Liu. Terahertz quantum cascade lasers oper-

ating up to 200 k with optimized oscillator strength and improved injection tunneling. *Opt. Express*, 20(4):3866–3876, 2012.

[10] Lord Rayleigh. On waves propagated along the plane surface of an elastic solid. *Proceedings of the London Mathematical Society*, 17:4–11, 1885.

[11] RM White and FW Voltmer. Direct piezoelectric coupling to surface elastic waves. *Applied Physics Letters*, 7(12):314–316, 1965.

[12] Ken-ya Hashimoto. *Surface acoustic wave devices in telecommunication: modelling and simulation*. Springer, 2000.

[13] S. Büyükköse, B. Vratzov, J. van der Veen, P. V. Santos, and W. G. van der Wiel. Ultrahigh-frequency surface acoustic wave generation for acoustic charge transport in silicon. *Applied Physics Letters*, 102(1):013112–013112–4, 2013.

[14] S Büyükköse, B Vratzov, D Ataç, J van der Veen, PV Santos, and WG van der Wiel. Ultrahigh-frequency surface acoustic wave transducers on zno/sio2/si using nanoimprint lithography. *Nanotechnology*, 23(31):315303, 2012.

[15] CK Campbell. *Surface acoustic wave devices for mobile and wireless communications*. Elsevier, 1998.

[16] J M Shilton, V I Talyanskii, M Pepper, D A Ritchie, JEF Frost, CJB Ford, CG Smith, and GAC Jones. High-frequency single-electron transport in a quasi-one-dimensional GaAs channel induced by surface acoustic waves. *JOURNAL OF PHYSICS-CONDENSED MATTER*, 8(38):L531–L539, SEP 16 1996.

[17] S H Simon. Coupling of surface acoustic waves to a two dimensional electron gas. *PHYSICAL REVIEW B*, 54(19):13878–13884, NOV 15 1996.

[18] GR Aizin, G Gumbs, and M Pepper. Screening of the surface-acoustic-wave potential by a metal gate and the quantization of the acoustoelectric current in a narrow channel. *PHYSICAL REVIEW B*, 58(16):10589–10596, OCT 15 1998.

[19] G Gumbs, GR Aizin, and M Pepper. Interaction of surface acoustic waves with a narrow electron channel in a piezoelectric material. *PHYSICAL REVIEW B*, 57(3):1654–1663, JAN 15 1998.

[20] AO Govorov, AV Kalameitsev, M Rotter, A Wixforth, JP Kotthaus, KH Hoffmann, and N Botkin. Nonlinear acoustoelectric transport in a two-dimensional electron system. *PHYSICAL REVIEW B*, 62(4):2659–2668, JUL 15 2000.

[21] R STONELEY. The Propagation of Surface Elastic Waves in a Cubic Crystal. *Proceedings of the Royal Society of London Series A-Mathematical and Physical Sciences*, 232(1191):447–458, 1955.

[22] E Cambiaggio and F Cuozzo. Finite difference analysis of surface acoustic wave propagation and scattering in piezoelectric crystals. *Journal of Computational Physics*, 33(2):153–168, 1979.

[23] S. Rahman, H. P. Langtangen, and C. H. W. Barnes. A finite element method for modelling electromechanical wave propagation in anisotropic piezoelectric media. *COMMUNICATIONS IN COMPUTATIONAL PHYSICS*, 2(2):271–292, APR 2007.

[24] Mikhail V. Kisin and Serge Luryi. Piezoelectric modulation of gain and distributed feedback for quantum cascade lasers with widely tunable emission wavelength. *Applied Physics Letters*, 82(6):847–849, 2003.

[25] Amnon Yariv. *Introduction to optical electronics*. Holt, Rinehart and Winston, Inc., New York, NY, 1976.

[26] J Faist, C Gmachl, F Capasso, C Sirtori, DL Sivco, JN Baillargeon, and AY Cho. Distributed feedback quantum cascade lasers. *APPLIED PHYSICS LETTERS*, 70(20):2670–2672, MAY 19 1997.

[27] Terence H Risby and Frank K Tittel. Current status of midinfrared quantum and interband cascade lasers for clinical breath analysis. *Optical Engineering*, 49(11):111123–111123, 2010.

[28] Robert F Curl, Federico Capasso, Claire Gmachl, Anatoliy A Kosterev, Barry McManus, Rafał Lewicki, Michael Pusharsky, Gerard Wysocki, and Frank K Tittel. Quantum cascade lasers in chemical physics. *Chemical Physics Letters*, 487(1):1–18, 2010.

[29] A Kosterev, G Wysocki, Y Bakhirkin, S So, R Lewicki, M Fraser, F Tittel, and RF Curl. Application of quantum cascade lasers to trace gas analysis. *Applied Physics B*, 90(2):165–176, 2008.

[30] Colette McDonagh, Conor S Burke, and Brian D MacCraith. Optical chemical sensors. *Chemical reviews*, 108(2):400–422, 2008.

[31] Matthew R McCurdy, Yury Bakhirkin, Gerard Wysocki, Rafal Lewicki, and Frank K Tittel. Recent advances of laser-spectroscopy-based techniques for applications in breath analysis. *Journal of Breath Research*, 1(1):014001, 2007.

[32] A Fried, B Henry, B Wert, S Sewell, and JR Drummond. Laboratory, ground-based, and airborne tunable diode laser systems: performance characteristics and applications in atmospheric studies. *Applied Physics B*, 67(3):317–330, 1998.

[33] WH Fan, A Burnett, PC Upadhya, J Cunningham, EH Linfield, and AG Davies. Far-infrared spectroscopic characterization of explosives for security applications using broadband terahertz time-domain spectroscopy. *Applied spectroscopy*, 61(6):638–643, 2007.

[34] G Wysocki, R Lewicki, RF Curl, FK Tittel, L Diehl, F Capasso, M Troccoli, G Hofler, D Bour, S Corzine, et al. Widely tunable mode-hop free external cavity quantum cascade lasers for high resolution spectroscopy and chemical sensing. *Applied Physics B*, 92(3):305–311, 2008.

[35] Benjamin S. Williams. Terahertz quantum-cascade lasers. *NATURE PHOTONICS*, 1(9):517–525, SEP 2007.

[36] Mohammed Salih. *Development of terahertz frequency quantum cascade lasers*. PhD thesis, University of Leeds, 2011.

[37] Mohammed Salih, Paul Dean, John Cunningham, Suraj P Khanna, Lianhe Li, Li Chen, A Giles Davies, and Edmund H Linfield. Terahertz quantum cascade lasers with angled facets for monolithic integration. In *Infrared Millimeter and Terahertz Waves (IRMMW-THz), 2010 35th International Conference on*, pages 1–2. IEEE, 2010.

[38] H Luo, SR Laframboise, ZR Wasilewski, GC Aers, HC Liu, and JC Cao. Terahertz quantum-cascade lasers based on a three-well active module. *Applied physics letters*, 90(4):041112–041112, 2007.

[39] M. Salih, P. Dean, A. Valavanis, S.P. Khanna, L.H. Li, J.E. Cunningham, A.G. Davies, and E.H. Linfield. Terahertz quantum cascade lasers with thin resonant-phonon depopulation active regions and surface-plasmon waveguides. *Journal of Applied Physics*, 113(11):113110–113110–5, Mar 2013.

[40] Safa O Kasap. *Optoelectronics & Photonics: Principles & Practices.* Pearson Higher Ed, 2012.

[41] Paul Harrison. *Quantum Wells, Wires and Dots.* Wiley, Chichester, 2nd edition, 2006.

[42] Stefan Birner, Tobias Zibold, Till Andlauer, Tillmann Kubis, Matthias Sabathil, Alex Trellakis, and Peter Vogl. Nextnano: General purpose 3-d simulations. *Electron Devices, IEEE Transactions on*, 54(9):2137–2142, 2007.

[43] Alexander Valavanis. *n-type silicon-germanium based terahertz quantum cascade lasers.* PhD thesis, University of Leeds, 2009.

[44] Leon James McBride Lever. *Charge transport in discotic liquid crystals.* PhD thesis, University of Leeds, 2006.

[45] Craig Anthony Evans. *The optical and thermal properties of quantum cascade lasers.* PhD thesis, University of Leeds, 2008.

[46] V. D. Jovanović, Z. Ikonić, D. Indjin, P. Harrison, V. Milanović, and R. A. Soref. Designing strain-balanced GaN/AlGaN quantum well structures: Application to intersubband devices at 1.3 and 1.55 μm wavelengths. *Journal of Applied Physics*, 93(6):3194–3197, 2003.

[47] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide.* Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition, 1999.

[48] D.F. Nelson, R.C. Miller, and D.A. Kleinman. Band Nonparabolicity Effects in Semiconductor Quantum-wells. *PHYSICAL REVIEW B*, 35(14):7770–7773, MAY 15 1987.

[49] Paul Harrison. *Computational Methods in Physics, Chemistry and Biology.* Wiley, 2001.

[50] Safa O Kasap. *Principles of electronic materials and devices*, volume 3. McGraw-Hill New York, NY, 2006.

[51] J. F. Nye and R. B. Lindsay. *Physical Properties of Crystals: Their Representation by Tensors and Matrices.* Clarendon Press, Oxford, 1957.

[52] J Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2):185–200, 1994.

[53] Z.S. Sacks, D.M. Kingsland, R. Lee, and Jin-Fa Lee. A perfectly matched anisotropic absorber for use as an absorbing boundary condition. *Antennas and Propagation, IEEE Transactions on*, 43(12):1460–1463, 1995.

[54] F. L. Teixeira and W. C. Chew. Complex space approach to perfectly matched layers: a review and some new developments. *International Journal of Numerical Modelling-Electronic Networks Devices and Fields*, 13(5):441–455, SEP-OCT 2000.

[55] W. C. Chew and Q. H. Liu. Perfectly matched layers for elastodynamics: A new absorbing boundary condition. *Journal of Computational Acoustics*, 4(4):341–359, DEC 1996.

[56] Steve G. Johnson. Notes on perfectly matched layers (pmls). May be found at http://math.mit.edu/ stevenj/18.369/pml.pdf, March 2010.

[57] P. M. Smith and R. Ren. Finite-difference time-domain techniques for SAW device analysis. In *Ultrasonics Symposium, 2002. Proceedings. 2002 IEEE*, volume 1, pages 325–328, oct. 2002.

[58] F. Chagla and P. M. Smith. Finite difference time domain methods for piezoelectric crystals. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 53(10):1895–1901, october 2006.

[59] A. Taflove and M. E. Brodwin. Numerical solution of steady-state electromagnetic scattering problems using the time-dependent Maxwell's equations. *Microwave Theory and Techniques, IEEE Transactions on*, 23(8):623–630, aug 1975.

[60] H. Schwerppe. Electromechanical properties of bismuth germanate $Bi_4(GeO_4)_3$. *Sonics and Ultrasonics, IEEE Transactions on*, 16(4):219, Oct. 1969.

[61] MR Buitelaar, PJ Leek, VI Talyanskii, CG Smith, D Anderson, GAC Jones, J Wei, and DH Cobden. Charge pumping and current quantization in surface acoustic-wave-driven carbon nanotube devices. *Semiconductor science and technology*, 21(11):S69, 2006.

[62] H Kogelnik and CV Shank. Coupled-wave theory of distributed feedback lasers. *Journal of applied Physics*, 43(5):2327–2335, 1972.

[63] Thomas G Jurgens, Allen Taflove, Korada Umashankar, and Thomas G Moore. Finite-difference time-domain modeling of curved surfaces [em scattering]. *Antennas and Propagation, IEEE Transactions on*, 40(4):357–366, 1992.

[64] Robert L Miller, Carl E Nothnick, and Dana S Bailey. *Acoustic charge transport: device technology and applications*. Artech House, 1992.

[65] Rodney Loudon. *The Quantum Theorie of Light*. Oxford University Press, New York, 1983.

[66] Scott Meyers. *Effective C++: 55 specific ways to improve your programs and designs*. Pearson Education, 2005.

[67] Thomas J Slight, Giuseppe Tandoi, Dmitry G Revin, Andrew McKee, Shiyong Y Zhang, Wyn Meredith, John W Cockburn, and Charles N Ironside. $\lambda \sim 3.35\,\mu$m distributed-feedback quantum-cascade lasers with high-aspect-ratio lateral grating. *Photonics Technology Letters, IEEE*, 23(7):420–422, 2011.

[68] MV Vasnetsov, V Yu Bazhenov, SS Slussarenko, and G Abbate. Coupled-wave analysis of second-order bragg diffraction. i. reflection-type phase gratings. *JOSA B*, 26(4):684–690, 2009.

[69] J Cunningham, VI Talyanskii, JM Shilton, M Pepper, MY Simmons, and DA Ritchie. Single-electron acoustic charge transport by two counterpropagating surface acoustic wave beams. *Physical Review B*, 60(7):4850, 1999.

[70] A Wixforth, JP Kotthaus, and G Weimann. Quantum oscillations in the surface-acoustic-wave attenuation caused by a two-dimensional electron system. *Physical review letters*, 56(19):2104, 1986.

[71] NE Fletcher, J Ebbecke, TJBM Janssen, FJ Ahlers, M Pepper, HE Beere, and DA Ritchie. Quantized acoustoelectric current transport through a static quantum dot using a surface acoustic wave. *Physical Review B*, 68(24):245310, 2003.

[72] CHW Barnes, JM Shilton, and AM Robinson. Quantum computation using electrons trapped by surface acoustic waves. *Physical Review B*, 62(12):8410, 2000.

[73] V Miseikis, JE Cunningham, K Saeed, R ORorke, and AG Davies. Acoustically induced current flow in graphene. *Applied Physics Letters*, 100(13):133105–133105, 2012.